Ecotoxicology and Environmental Safety 289 (2025) 117640

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Metagenomic exploration and computational prediction of novel enzymes for **polyethylene terephthalate** degradation

Donya Afshar Jahanshahi [a,b], Mohammad Reza Rezaei Barzani [b], Mohammad Bahram [c,d,e], Shohreh Ariaeenejad [f,*], Kaveh Kavousi [b,*]

[a] Department of Bioinformatics, Kish International Campus University of Tehran, Kish, Iran
[b] Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran
[c] Department of Ecology, Swedish University of Agricultural Sciences, Uppsala 75007, Sweden
[d] Institute of Ecology and Earth Sciences, University of Tartu, 40 Lai St., 51005, Tartu, Estonia
[e] Department of Agroecology, Aarhus University, Forsøgsvej 1 4200, Slagelse, Denmark
[f] Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research Education and Extension Organization (AREEO), Karaj, Iran

ARTICLE INFO

Edited by Dr R Pereira

Keywords:
Plastic-contaminated soil
Metagenomics
Plastizymes
PET degradation
machine learning
High-throughput screening

ABSTRACT

As a global environmental challenge, plastic pollution raises serious ecological and health concerns owing to the excessive accumulation of plastic waste, which disrupts ecosystems, harms wildlife, and threatens human health. Polyethylene terephthalate (PET), one of the most commonly used plastics, has contributed significantly to this growing crisis. This study offers a solution for plastic pollution by identifying novel PET-degrading enzymes. Using a combined approach of computational analysis and metagenomic workflow, we identified a diverse array of genes and enzymes linked to plastic degradation. Our study identified 1305,282 unmapped genes, 36,000 CAZymes, and 317 plastizymes in the soil samples were heavily contaminated with plastic. We extended our approach by training machine learning models to discover candidate PET-degrading enzymes. To overcome the scarcity of known PET-degrading enzymes, we used a Generative Adversarial Network (GAN) model for dataset augmentation and a pretrained deep Evolutionary Scale Language Model (ESM) to generate sequence embeddings for classification. Finally, 21 novel PET-degrading enzymes were identified. These enzymes were further validated through active site analysis, amino acid composition analysis, and 3D structure comparison. Additionally, we isolated bacterial strains from contaminated soils and extracted plastizymes to demonstrate their potential for environmental remediation. This study highlights the importance of biotechnological solutions for plastic pollution, emphasizing scalable, cost-effective processes and the integration of computational and metagenomic methods.

1. Introduction

The world's annual production of plastics, integral to the global economy, exceeds 400 million tons (Pathak and Navneet, 2017; Lear et al., 2021). However, environmental repercussions are severe, as slow degradation rates result in the accumulation of 6.3 billion metric tons of global plastic waste. Microplastics have become an increasing environmental hazard that threatens various life forms, including humans.

Abbreviation: GAN, Generative Adversarial Network; ESM, Evolutionary Scale Modeling; SVM, support vector machine; RF, random forest; KNN, k-nearest neighbor; MAGs, metagenome assemble genome; GH, glycoside hydrolases; AA, auxiliary activity; CBM, carbohydrate-binding modules; GT, Glycosyl transferase; CE, Carbohydrate Esterase; PL, Polysaccharide Lyase; CPL, caprolactam; PE, Polyethylene; PET, Polyethylene terephthalate; PA, Polyamide; PVC, Polyvinyl chloride; PUR, polyurethane; Asp, Aspartic acid; His, Histidine; Ser, Serine; Cys, cysteine; Asp, aspartate; Ala, alanine; Arg, arginine; Glu, glutamate; Pro, proline; Asn, asparagine; Phe, phenylalanine; CAZymes, carbohydrate-active enzymes; Plastizymes, Plastic-degrading enzymes; PLDDT, predicted local distance difference test scores, e value, expectation value; COG, clusters of orthologous groups; MeTarEnz, metagenomic targeted enzyme miner; SRA, Sequence Read Archive.

* Correspondence to: Laboratory of Complex Biological Systems and Bioinformatics (CBB), Department of Bioinformatics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, P.O. Box: 13145-1384, 1417614411, Iran.
E-mail addresses: sh.ariaee@abrii.ac.ir, shariaee@gmail.com (S. Ariaeenejad), kkavousi@ut.ac.ir (K. Kavousi).

https://doi.org/10.1016/j.ecoenv.2024.117640
Received 29 June 2024; Received in revised form 13 December 2024; Accepted 29 December 2024
Available online 10 January 2025

Current plastic waste management methods are far from efficient, highlighting the need for innovative solutions. To address this issue, two main approaches have been explored for pollutant degradation: chemical catalysts, such as nanomaterials (Hashim Khudhair et al., 2023; Yousefzadeh et al., 2023), and biological catalysts, like enzymes. Both chemical catalysts and enzyme-based biodegradation offer effective approaches to degrading organic pollutants, providing valuable solutions for pollution treatment. In this regard, harnessing microbes for plastic biodegradation offers a promising approach (Réjasse et al., 2022; Kawai et al., 2019; Palm et al., 2019; Sowmya et al., 2015; Tao et al., 2023).

Isolating plastic-degrading enzymes such as PETase and cutinase (Charnock, 2021) remains a challenge, primarily due to the cultivation difficulties associated with many microorganisms (Jahanshahi et al., 2023). Metagenomics offers a promising approach for comprehensive exploration of diverse microorganisms involved in plastic degradation. Using advanced computational approaches, such as machine learning allow the prediction of various enzyme properties based on sequence similarity e.g. (Ariaeenejad et al., 2022b; Foroozandeh Shahraki et al., 2021, 2020; Shahraki and F.F.A.S.A.M.R.G.M.H.N.-B.M.M.G.H.S.K.K., 2022; Ariaeenejad et al., 2023) Previous studies have successfully identified enzymes capable of degrading various plastics.

The degradation of PET involves enzymes, primarily PETases and esterases, which are produced by microorganisms such as *Ideonella sakaiensis*. PETase contains key amino acids (Austin et al., 2018; Joo et al., n.d.; Liu et al., 2018; Samak et al., 2020). The analysis of the PETase enzyme's structure has identified three crucial amino acids, namely Serine (Ser160), Histidine (His237), and Aspartic acid (Asp206), located in the active cleft. Further studies have shown that these three amino acids are key components of the cutinase enzymes involved in PET degradation (Kitadokoro et al., 2019).

To tackle the growing plastic waste crisis, recent studies highlight the transformative role of artificial intelligence (AI) in discovering and optimizing plastic-degrading enzymes. Mican, Jaradat, and colleagues introduced a workflow integrating in-silico and high-throughput approaches, enabling the rapid identification of PET-degrading enzymes (Mican et al., 2024). Similarly, Jiang et al. developed PEZy-Miner, a machine-learning framework using protein language models (pLMs) to analyze enzyme sequences, identifying 27 high-confidence candidates from over 0.1 million sequences(“TY - JOUR AU - Jiang, Renjing AU - Yue, Zhenrui AU - Shang, Lanyu AU - Wang, Dong AU - Wei, Na PY.).

Gupta and Agrawal advanced PETase optimization through machine-learning-guided directed evolution, achieving improved enzyme efficiency with higher optimal degradation temperatures (up to 71.38°C) (Gupta and Agrawal, 2023). Wu et al. combined protein language models and structural analysis to identify 14 robust PET hydrolases, including KbPETase, which surpasses existing PETases in efficiency and stability. These studies demonstrate the potential of AI-driven methods to accelerate enzyme discovery and provide sustainable solutions for plastic recycling (Wu et al., 2024).

This study aimed to explore the potential of soil microbes to discover enzymes capable of effectively degrading plastics, by leveraging advanced computational tools to uncover new candidates in heavily contaminated soils. A classification pipeline was developed to identify these enzymes in metagenomes. Using the ESM-2 (“Zeming Lin et al.) protein language model, we generated embeddings from protein sequences to capture their structural and chemical properties as input for the classification process. Given the scarcity of confirmed PET degrading enzymes, we employed Generative Adversarial Networks (GANs) for dataset augmentation, generating similar sequences to create a larger training dataset for the classifiers. To validate the PET degradation potential of the candidate sequences, we thoroughly analyzed their active sites and compared the key amino acid residues with those identified in previously reported, experimentally validated PET-degrading enzymes. We further conducted a wet-lab experiment to validate the effectiveness of the PET-degrading enzymes under specific environmental conditions. Our results demonstrated the degradation of Low-density Polyethylene (LDPE) by PET-degrading enzymes. It was thoroughly investigated using advanced characterization techniques including Atomic Force Microscopy (AFM), Fourier Transform Infrared Spectroscopy (FTIR), and Scanning Electron Microscopy (SEM). These methods provide a comprehensive understanding of the degradation process at molecular and structural levels. By uniquely combining machine learning predictions, active site validation, and wet-lab experimentation, we introduced a more robust and comprehensive approach compared to previous studies.

2. Material and methods

2.1. Collection of samples and extraction of metagenomic DNA

The arrangement of sampling points for DNA collection depends on the study objectives and type of environment.

In this study, ten sampling points were selected randomly within agricultural areas in Pishva and Varamin, Tehran Province, with geographical coordinates of (35.1848° N latitude, 51.4214° E) and (35.465° N latitude, 51.594° E), respectively, to minimize bias. The land in question was under mulch cultivation and irrigated with wastewater. These soils have been contaminated with plastic residues, particularly polyethylene (PE), polyethylene terephthalate (PET), and nylon, for approximately 35 years. Soil samples were collected at depths ranging from 5 to 20 cm, a depth range often chosen in metagenomic studies because it represents the biologically active zone, including the rhizosphere and the surrounding soil layers. The upper 5–30 cm of soil typically contains higher concentrations of organic matter, nutrients, and microbes, supporting a diverse microbial community, including bacteria, fungi, and archaea. The surface soil (0–5 cm) is more influenced by external factors, such as sunlight and transient organic matter, which might not accurately reflect the stable soil microbial community. The 5–20 cm depth range includes both aerobic and anaerobic zones, providing a more comprehensive view of the microbial diversity (“Qianqian Qin, Yanhong Liu.), (Nutaratat et al., 2024).

The collected samples were stored in glass containers on dry ice to maintain the required temperature. Samples were pooled in the laboratory before DNA extraction and sequencing. DNA was extracted from a membrane filter (cellulose ester, Millipore, Billerica, MA, United States) using a FastDNA Spin Kit (MP Biomedicals, Solon, OH, United States). The quality and quantity of the extracted DNA were assessed using agarose gel electrophoresis and a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA).” (Ariaeenejad et al., 2022a; Ariaeenejad et al., 2024).

2.2. Metagenome library preparation and sequencing

Illumina sequencing libraries were prepared using TruSeq DNA Library Preparation Kit v2 (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. A Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA) was used to determine the quantity of each sample. All libraries were sequenced by Novogene Inc. (Beijing, China) using Illumina HiSeq 2500 sequencing system.

2.3. Taxonomic profiling

To ensure the accuracy of the raw reads, FastQC software was employed for quality control. In order to obtain more comprehensive taxonomical profiling results, two distinct methods were used. Prior to assembly and the creation of contigs, MetaPhlAn3 was employed for taxonomic profiling of the sample. Furthermore, after the assembly process and generation of contigs and genome bins, GTDB-tk was utilized to assign taxonomy to the metagenome assemble genomes (MAGs).

2.4. Functional profiling

2.4.1. Bioinformatics analysis for gene and enzymatic groups prediction

The raw reads obtained from sequencing were analyzed using MEGAHIT to assemble high-quality short reads. Two methods, single-assembly and mixed-assembly, were employed during the assembly process to enhance the binning quality in subsequent stages (Fig. 1). Subsequently, the paired-end reads were mapped to the resulting contigs using BWA to determine their coverage profiles. Genome bins were reconstructed using MetaBAT2 software. The bins were merged. The duplicated bins were reduced using dRep software. Completeness, contamination, and strain heterogeneity of the genome bins were assessed using the CheckM program. MetaGenMark was used to identify protein-coding regions within the metagenome sequences of each bin. CD-HIT was employed to ensure the uniqueness of the gene. Additionally, KofamKOALA was utilized to assign the non-redundant gene catalog to KEGG orthology and eggNOG-mapper was used to allocate functional categories within clusters of orthologous groups (COG). Furthermore, all generated bins were analyzed using ECpred to detect potential enzymatic families. In addition, standalone run_dbCAN2 was applied to identify CAZyme genes. This workflow enabled the analysis of a metagenome sample from plastic-contaminated soil to generate corresponding genes, taxa information, and novel plastizymes identification (Fig. 1).

2.4.2. Identifying of the novel plastizymes

A comprehensive dataset encompassing plastic-degrading enzymes was gathered using the NCBI protein databases, BRENDA, and UniProt. The compilation was derived from an extensive literature review and encompasses experimentally validated plastizymes. To redundancy reduction, CD-HIT was employed to cluster highly homologous sequences with a cut-off value of 0.9, resulting in one representative sample being retained from each cluster. Ultimately, 158 sequences were selected for the final dataset, each of which had experimental evidence for plastic degradation. Both binned and unbinned samples were examined by the MeTarEnz tool (Mehdi F. Shahraki, 2022) to identify potential plastic-degrading genes. The minimum bit-score cut-off was set to 250. The resulting metagenomic sequences were subsequently analyzed using NCBI CDD, AlphaFold2 and TM-align for 3D structure prediction and comparison for all putative plastizymes for further investigation.

2.4.3. Advanced analysis for the identification of novel PET-degrading enzymes

This study focused exclusively on the enzymes that degrade PET. By utilizing MeTarEnz (Shahraki and F.F.A.S.A.M.R.G.M.H.N.-B.M.M.G.H. S.K.K., 2022) and NCBI CDD to evaluate factors such as bit-score, expectation value (e-value), and superfamily similarity, 32 enzymes with the potential to break down PET were identified. Next, a comprehensive machine-learning model was employed to predict PET-degrading enzymes more accurately. To do so, it is crucial to have an appropriate amount of data to effectively train the classifiers. Because there was an insufficient amount of PET-degrading dataset available, a Generative Adversarial Network (GAN) was utilized for data augmentation (Lan et al., 2020;Liu et al., 2019). Protein GAN, on the other hand, refers to the application of GANs specifically for generating synthetic protein sequences with the desired properties (Repecka et al., 2021).

The training dataset comprised 120 sequences of cutinase and PETase enzymes acquired from the NCBI and UniProt databases. To evaluate the generated sequences, they were aligned with the training and validation datasets using BLAST at intervals of 1200 training steps. The training process was conducted on a GPU P100 with 16 GB of

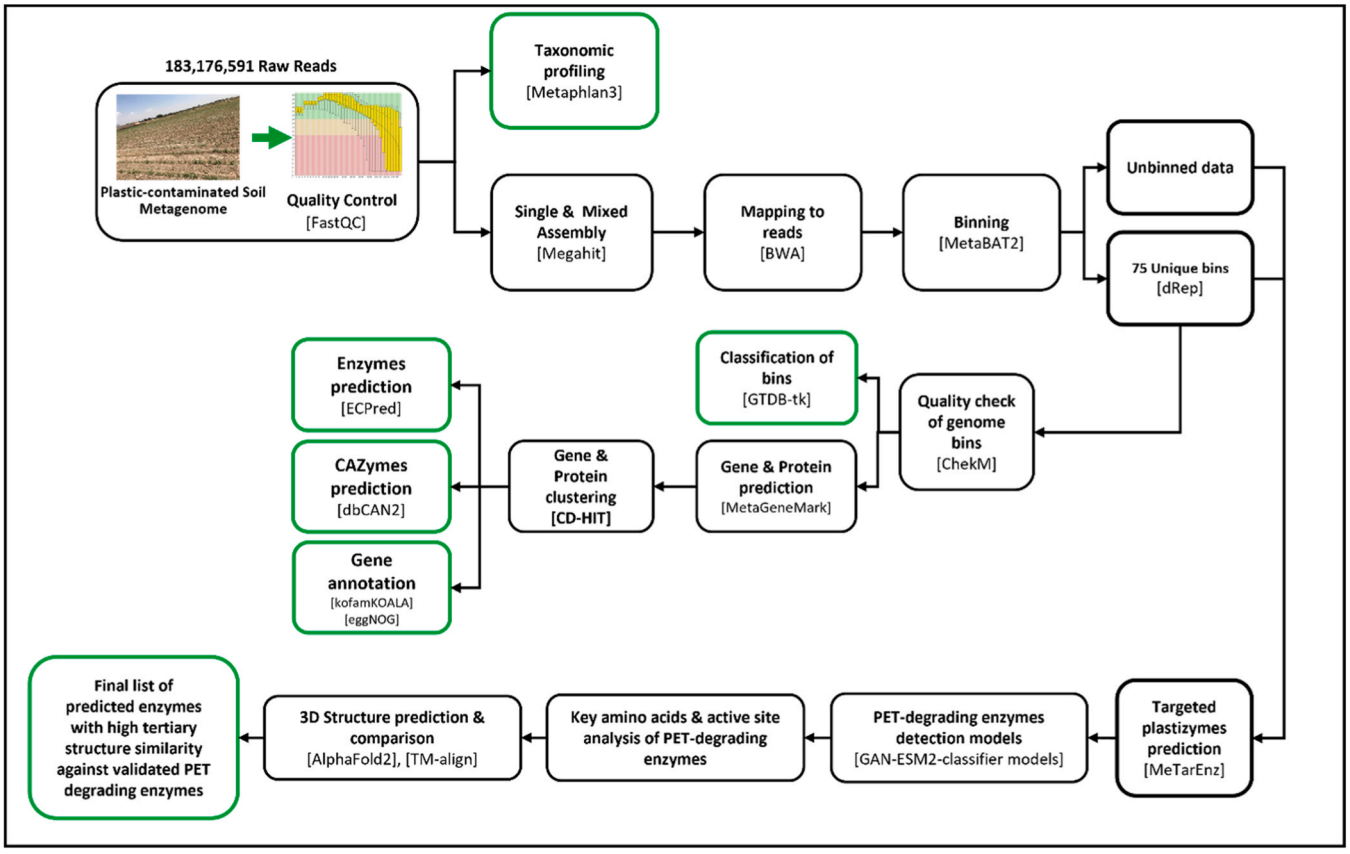

Fig. 1. The pipeline utilized to process the raw metagenome samples for the identification of Bins, Genes, taxa, Enzymes, CAZYmes, and putative targeted plastizymes.

memory, taking approximately 9.3 hours to complete.

To reduce data complexity and extract features, a language model (LM) based on ESM-2 (Lin et al., 2023) was employed to generate an embedded matrix. ESM-2 is another example of a machine-learning method that has been successfully applied in the field of bioinformatics. ESM-2 is a state-of-the-art deep learning LM that uses a transformer-based neural network architecture to generate highly accurate predictions of protein structures.

A pretrained ESM-2 LM was employed. Training and test data were categorized into two groups. The positive class consisted of 204 PET-degrading enzymes, which included both confirmed enzymes and synthetic sequences generated by the Protein GAN. The negative class comprised 130 cutinase enzymes that did not degrade PET.

Negative data were collected using expert knowledge. The active sites and critical amino acids involved in PET degradation were analyzed. Sequences that lacked essential features for plastic degradation and binding to PET, were categorized as negative class.

The ESM-2 model generates an embedding matrix (256 x 256) for each sequence. Finally, the embedded matrix served as input for the training support vector machine (SVM), random forest (RF), and k-nearest neighbor (KNN) classifiers. These classifiers were utilized to classify candidate enzymes selected by MeTarEnz from the metagenome sample (Ariaeenejad et al., 2022a; Motamedi et al., 2021). The detailed workflow of this stage is shown in Fig. 2.

After that the putative PET-degrading enzymes were identified using the prediction model, the presence of key amino acids and active site composition and 3D structure comparison using AlphaFold2 and TM-align were analyzed to enhance the validation of putative enzyme sequences.

2.5. Experimental validation

2.5.1. Isolation of plastic-degrading bacteria

This study aimed to isolate plastic-degrading bacteria, specifically targeting those with cutinases, PETases, and laccases. To achieve this, 1 g each of Cutin and PE substrates were added to 50 ml of synthetic medium (SM). This media comprised various components: Substrate (10.0 g), NH_4NO_3 (5.0 g), $MgSO_4.7H_2O$ (0.5 g), KH_2PO_4 (1.0 g), $CaCl_2.2H_2O$ (0.148 g), KCl (0.5 g), NaCl (0.03) and trace elements (0.01 g of $FeSO_4.7H_2O$ and 0.008 g $MnSO_4$) in 1000 ml of distilled water. In addition, 1 g of plastic-contaminated soil was added to each flask. The resultant enrichment medium was then incubated at 30°C for 1 days on a rotary shaker set at 150 rpm.

After preparation, samples were spread on solid medium plates of SM and nutrient agar, and then incubated at 30°C for three days to facilitate bacterial growth. This period allowed for the development of bacterial colonies on mediums enriched with specific substrates for the two types of enzymes under controlled temperature and pH conditions. Following adequate growth, bacterial cells were harvested, usually through centrifugation, to isolate them from the growth medium containing enzyme extracts. These extracted enzymes were stored under conditions that preserved their stability and activity.

2.5.2. Biodegradation of low-density polyethylene by plastic-degrading enzymes

In the experiments, Low-density Polyethylene (LDPE) plastic carrier bags with a thickness of 20 μm were cut into 5 × 5 cm films. These films were thoroughly cleaned with distilled water and 100 % ethanol and then dried with filter paper. The sterilized films were treated with a mix of three plastic-degrading enzymes and incubated at 37 °C for 30 days, with 16 h of daily UV light exposure. After incubation, the films were removed, washed with distilled water and 75 % ethanol, and dried with filter paper (Li et al., 2020). The surface microstructures of these treated films were examined using Scanning Electron Microscope (SEM) (QUANTA 200, USA) at an acceleration voltage of 16 kV. Fourier Transform Infrared Spectroscopy with Attenuated Total Reflectance (FTIR-ATR) analysis was conducted in the frequency range of 4000–650 cm^{-1} using a PerkinElmer 100 FTIR instrument equipped with an ATR attachment in total reflection mode (%T). Furthermore, Atomic Force Microscopy (AFM) was employed to analyse the dried polyethylene films, capturing images at a scan speed of 1.0 Hz with a resolution of 200 × 200 pixels.

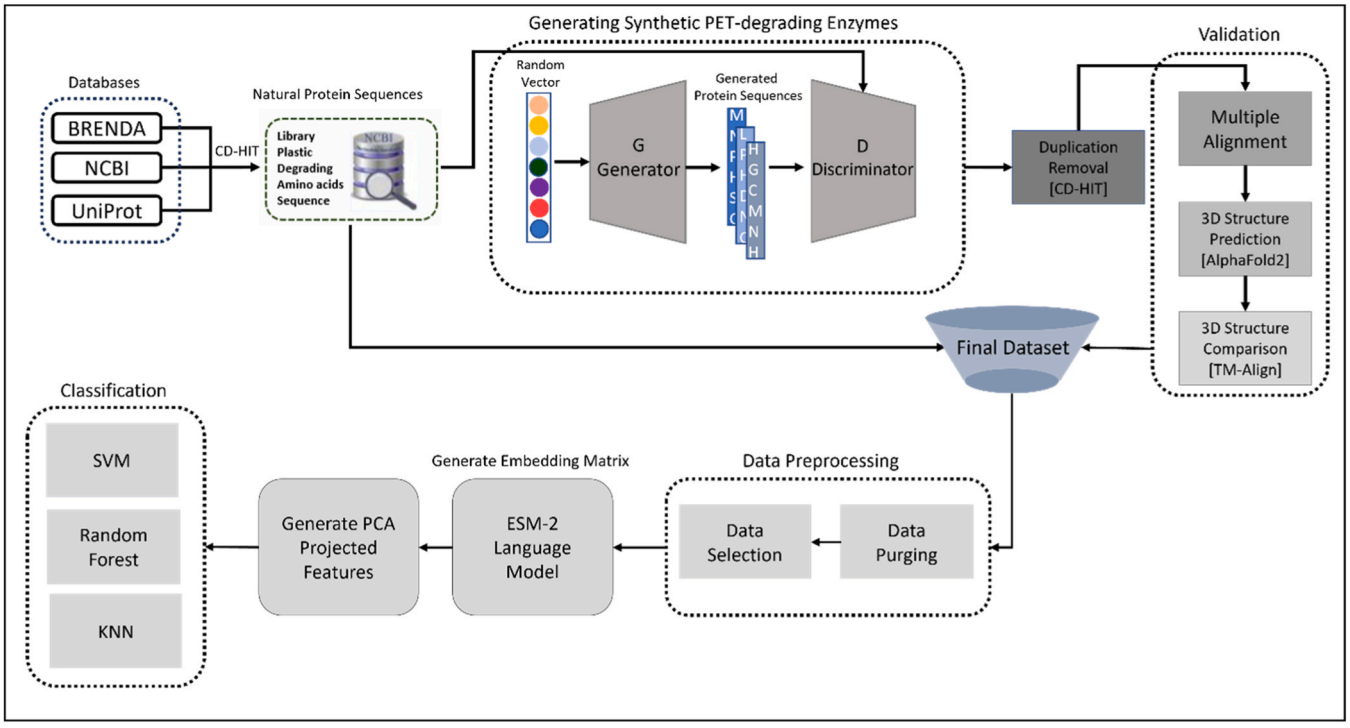

Fig. 2. The two-step workflow for generating artificial PET-degrading enzymes and training a model for classification and identification of novel PET-degrading enzymes.

3. Results and discussion

3.1. Microbial community structure

While microbial diversity and their functions in plastic-contaminated environments have not been extensively investigated, some recent studies have highlighted the potential of certain microorganisms and enzymes to degrade plastics and convert them into their monomers [1,2,64]. Notable examples include *Ideonella sakaiensis, Thermobifida fusca, Pseudomonas soli, Pseudomonas jessenii, Fusarium redolens* and, *Penicillium.*

Analysis of the raw reads revealed diverse microbiome in the sample. The most dominant phyla identified were *Proteobacteria* (69.78 %) and *Actinobacteria* (17.53 %). In addition, MetaPhlAn3 predicted a total of 10 classes, 21 orders, 38 families, 37 genera, and 224 non-redundant species in our sample. (Fig. S1A, B). Among these, we found four PET-degrading species (*Thermobifida fusca, Pseudomonas soli, Ideonella sakaiensis, and Pseudomonas jessenii)* within the top 28 species identified through taxonomic annotation.

75 individual genome bins were identified, of which 53 were found to have a completeness score above 65 % and a contamination score below 10 %. These individual genomes consist of bacteria and archaea. Furthermore, GTDB-tk compared bins to a 16S rRNA database and detected additional 18 archaeal genomes. In addition, three species of fungi, (*Chrysosporium sp, Aspergillus niger,* and *Fusarium solani*) (Ekanayaka et al., 2022) were observed within the collected sample which were reported to be PET-degrading microorganisms,.

Several bacterial species, including *Pseudomonas* (Nylon (Prijambada et al., 1995) and caprolactam (Otzen et al., 2018) digesters), *Thermobifida fusca* and *Ideonella sakaiensis* (PET digester), *Novosphingobium* (Phenanthrene (Rodriguez-Conde et al., 2016) digester) were reported as plastic degrading-bacteria. In our taxonomic annotation results, *Pseudomonas soli* was the most abundant plastic-degrading bacteria, constituting 3 % of all identified species. On the other hand, *Ideonella sakaiensis* accounted for only 0.4 % of the total species identified in the sample. The taxonomical profiling of plastic-degrading species is summarized in (Fig. 3), which displays a phylogenetic tree of a group of MAGs containing plastic-degrading bacteria. Through metagenomic analysis, we found that these species have evolved mechanisms to break down plastics because some of the identified plastizymes were found in their genomes.

3.2. Functional analysis of metagenomic sequences

The sequencing of the sample produced 183,176,591 high-quality reads. Functional analysis including enzyme prediction and plastizyme identification from both binned and unbinned data.

The assembly of short reads yielded 3.4 million contigs, with the longest at 600,358 bp (Fig. S2A) and identified 1.3 million unmapped genes/proteins. The enzyme family analysis is shown in (Fig. S2B) and highlighted oxidoreductases, hydrolases, and lyases, including plastizymes. This resulted in 157,344 genes mapped to KEGG orthologs and 670,772 genes annotated with COG categories via the Evolutionary Genealogy of Genes Non-supervised Orthologous Groups (eggnog) mapper (Fig. S2C).

We analyzed the Varamin soil sample using dbCAN2 to profile CAZymes. This revealed a total of 35,896 CAZyme-encoding genes in the metagenome sample, distributed among 51 CAZyme subclasses. Glycoside hydrolase (GH), a key enzyme involved in glycosidic bond breakdown, emerged as the most abundant CAZyme. It was closely followed by glycosyltransferase (GT), which plays a crucial role in glycosidic bond synthesis. Additionally, the presence of carbohydrate-binding

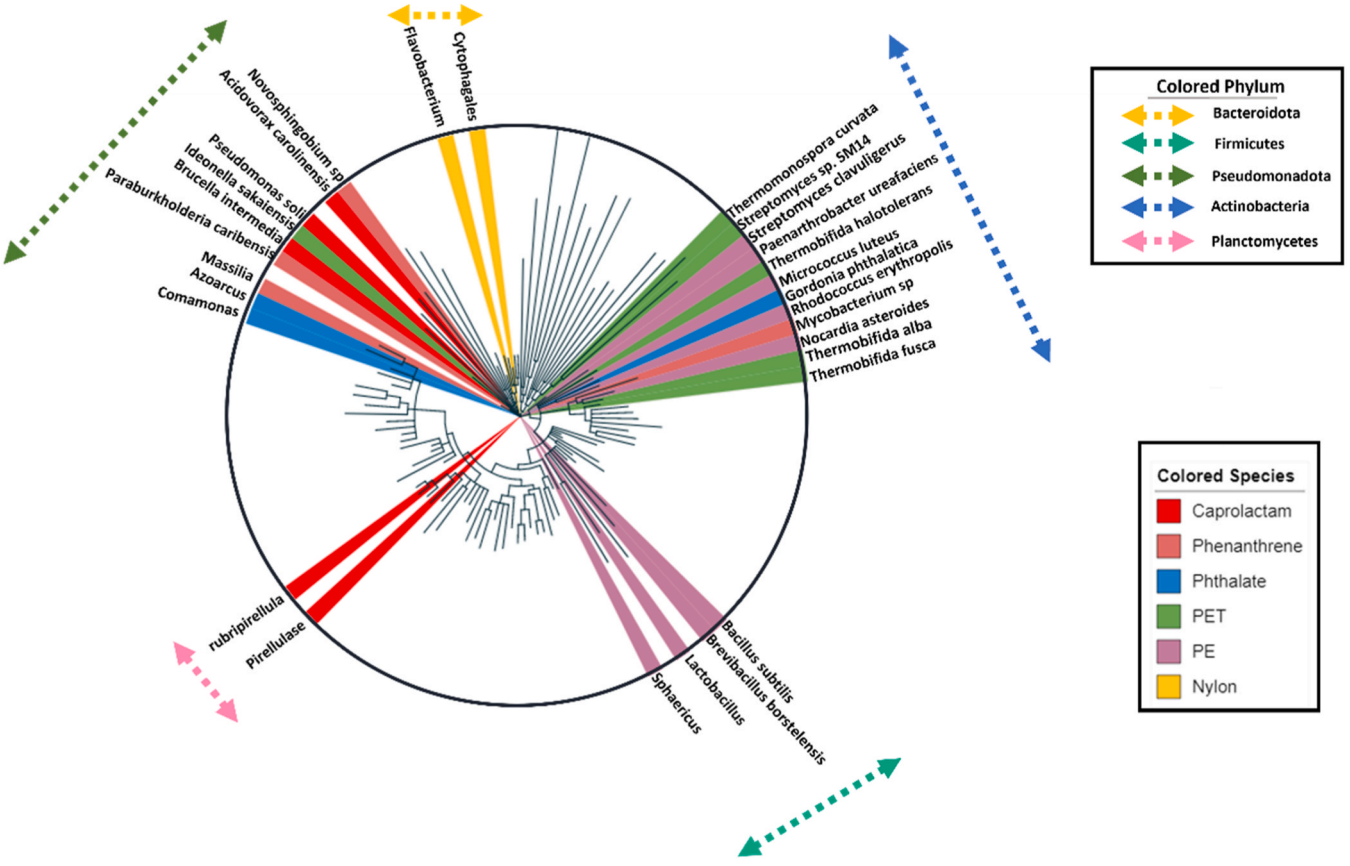

Fig. 3. Phylogenetic tree of a group of MAGs containing plastic-degrading bacteria. The dotted arrows show the phylum while the radial colors represent the type of plastic which the species could digest. Note: The taxonomic phylogenetic tree was visualized using Interactive Tree of Life (iTOL) v5 (https://itol.embl.de/).

modules (CBMs), responsible for guiding other CAZymes to their specific carbohydrate substrates, and Auxiliary Activities were particularly significant. Specifically, we found high relative abundances of GH1, CBM, AA1, AA3, and GT2 (Fig. S2D). These are families of CAZymes known to encompass numerous plastizymes. GH was the most abundant class, followed by GT and CBM (Fig. S2E).

3.2.1. *Metagenomic analysis for identifying of the plastizymes*

To further study plastizyme characteristics, the coverage of contigs against raw reads was analyzed using BWA alongside various tools to determine the ideal number of plastizymes matching source enzymes. Analysis of both binned and unbinned data with MeTarEnz (Shahraki and F.F.A.S.A.M.R.G.M.H.N.-B.M.M.G.H.S.K.K., 2022) revealed 3498 plastizymes in unbinned contigs and 3659 plastizymes in binned

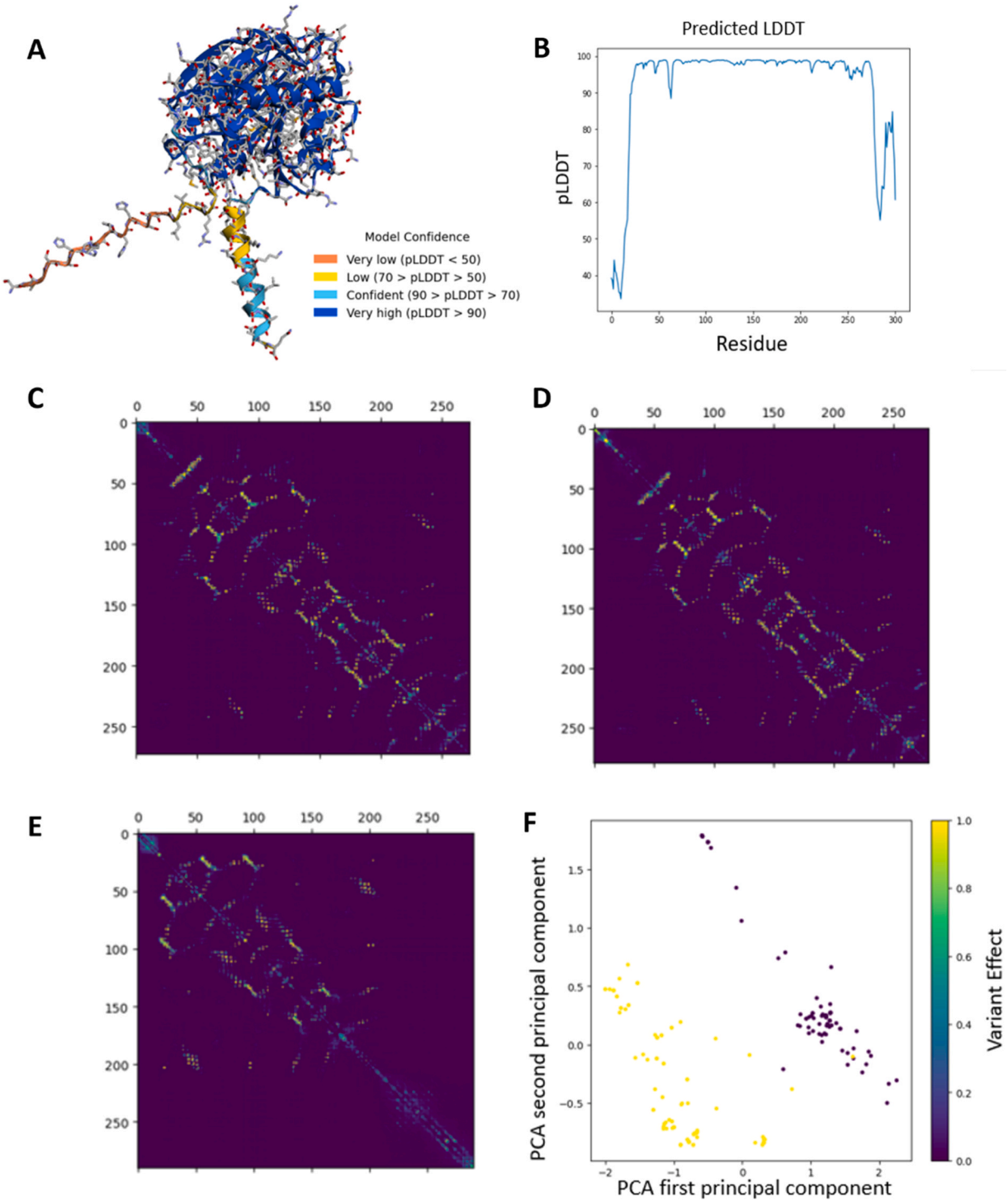

Fig. 4. Results of Protein GAN and ESM model. **A** 3D structure of PersiPETase1 that predicted by Alphafold2. **B** pLDDT (predicted local distance difference test) on each residue. **C** The contact map of one of the PET-degrading enzymes. **D** The contact map of another PET-degrading enzyme differs from the sequence in section C. **E** The contact map of one of the non-PET-degrading enzymes. Note: The horizontal and vertical axes represent the sequence positions of the amino acid residues. Each position is represented by a square and the color of the square indicates the degree of contact or interaction between the corresponding residues. **F** The PCA plot of PET and non-PET-degrading enzymes classification.

samples. Among the identified plastizymes.

317 enzymes were selected for high bit scores, low E-values, and strong plastic degradation potential, including PETase and laccase. These enzymes, covering 12 families, showed high pLDDT (>75 %) and TM-scores (59 %-98 %), indicating structural similarity and degradation affinity (Table S1).

The six most prevalent plastizymes identified in our metagenomic environment were proxidase, dehydrogenase, cutinase, PETase, hydrolase, and laccase (Fig. S3A). These enzymes play a pivotal role in the degradation of Caprolactam, Nylon, PET, and PE. Cutinase exhibited the greatest similarity to putative plastizymes (bit score: ~1600), followed by deoxygenase (bit score: ~1500) and 3-hydroxyacyl-CoA dehydrogenase (bit score: ~1300) (Fig. S3B, Fig. S3C). The relative abundance of plastizymes in our sample suggests that plastic-contaminated environments serve as reservoirs for plastic-degrading enzymes.

3.2.2. Enhanced analysis of novel PET-degrading enzymes through developing a customized prediction model

PET-degrading enzymes, notably PETases and Cutinases, are well studied plastizymes. A database of natural PET-degrading enzymes was created based on the literature review. Owing to the small size of the original dataset, ProteinGAN was used to augment the data and achieve high-quality results in data enhancement for classifier training. We found 100 PET-degrading enzyme sequences, which were validated through comparison with existing protein databases using tools such as CDD-NCBI. For further validation of the generated sequences, the 3D structure was predicted using AlphaFold, which was similar to the real sequences (Fig. 4A, B). This indicated that the predicted structure had a high validation score and was likely to be accurate.

Next, we generated an embedding matrix using the ESM-2 LM and applied three different classifiers to classify novel PET-degrading enzymes. This step allowed us to analyze and categorize the enzymes based on their specific properties and identify those with the highest potential for PET degradation. Further, ESM-2 representations were used to train three different contact maps of PET-degrading enzymes (Figs. 4C and 4D) compared to that of a non-PET-degrading enzyme (Fig. 4E). The contact maps of the first two enzymes were similar, indicating a common pattern of interactions between the amino acid residues involved in PET degradation. In contrast, the contact map of the non-PET-degradation enzyme was different, indicating a distinct pattern of interactions that did not support PET degradation.

We performed principal component analysis (PCA) to classify PET and non-PET-degrading enzymes based on their properties (Fig. 4F).

We then utilized the extracted representations to train the three models: SVM, KNN, and RF. The results of the evaluation showed that the RF model achieved the lowest error rate of 0.09, followed by the

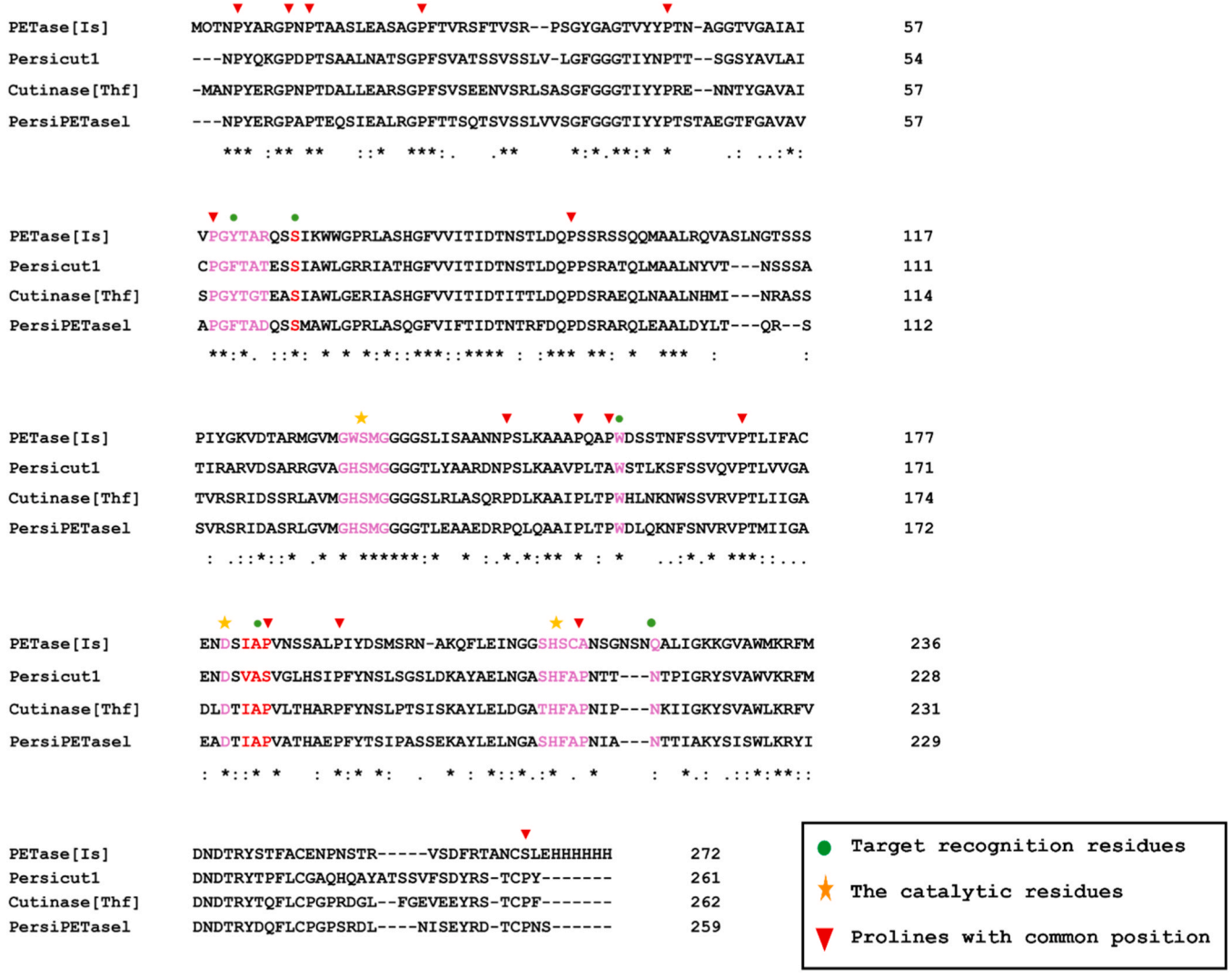

Fig. 5. Sequence alignment of PETase, Persicut1, Cutinase, and PersiPETase1 (The catalytic triad residues are marked with orange asterisks, and target recognition residues are marked with green circles. In addition, purple and red amino acids are key amino acids.).

SVM model with 0.1, and the KNN model with 0.1. In the final step, the candidate PET-degrading enzymes that had passed all previous validation steps were introduced to the models, and the models were used to successfully classify 21 sequences (PETase and cutinase) as the final candidate PET-degrading enzymes.

The amino acid residues that participate in the PET decomposition process are highly conserved within the proven and novel PET-degrading enzymes (Fig. 5). From the candidate enzymes screened by MeTarEnz (Shahraki and F.F.A.S.A.M.R.G.M.H.N.-B.M.M.G.H.S.K.K., 2022), NCBI CDD, PET degrading machine detection level, and amino acid composition screening, only 21 potential PET-degrading enzymes, including PETase and cutinase, were validated by AlphaFold 2 and TM-align tools with a confidence score of 100 %. We designated these innovative enzymes as PersiCuts and PersiPETases. The high similarity of the amino acid sequences of PersiCut1-PersiCut12 and PersiPETase1-PersiPETase9 with cutinase and PETase superfamily

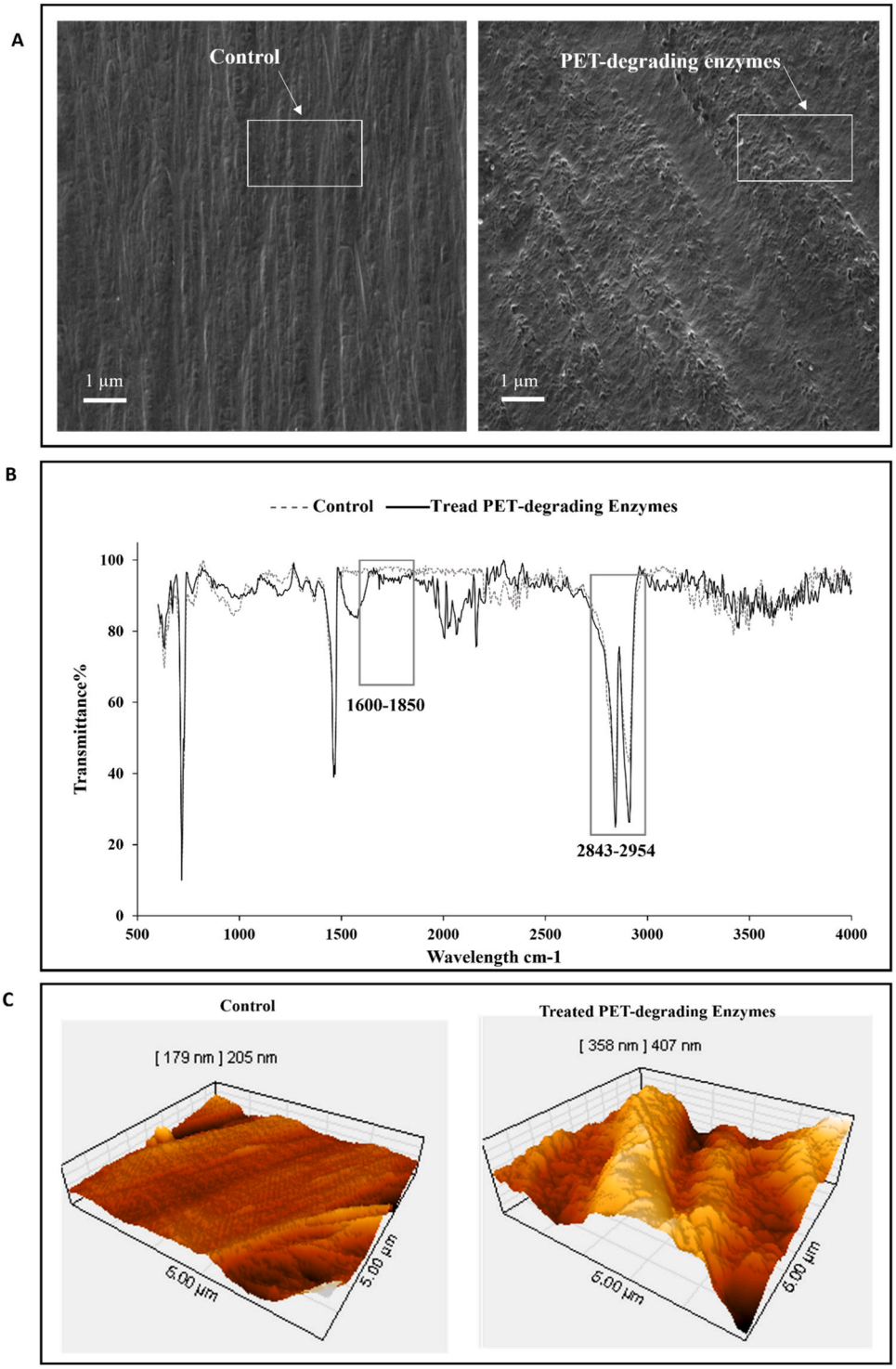

Fig. 6. Comparative analysis of LDPE films: SEM, FTIR, and AFM studies on the impact of plastic-degrading enzymes. **A** SEM images of LDPE films modified by plastic-degrading enzymes compared with control. **B** presents the FTIR spectra of LDPE films modified by plastic -degrading enzymes in comparison with control samples. **C** AFM of LDPE films modified by plastic -degrading enzymes compared with control.

members was confirmed by CDD. Fig. S4A illustrates the results of the 3D structure of AlphaFold prediction with the pLDDT (per residue Local Distance Difference Test) score (Fig. S4B) and the PAE (Predicted Aligned Error (Fig. S4C)) which had the highest model validation scores. A phylogenetic tree of the novel and proven PET- degradation enzymes from the collected database (Fig. S5) demonstrated the similarity of the experimentally introduced PET-degrading enzymes with our novel enzymes.

3.3. Efficiency assessment of plastic-degrading enzymes

The application of plastic-degrading enzymes derived from specific bacteria represents a transformative approach to address the challenge of polyethylene degradation. By catalyzing the depolymerization and oxidation processes, these enzymes convert the long, recalcitrant chains of polyethylene into smaller, environmentally benign molecules, accelerating degradation and offering a sustainable solution for recycling or decomposing polyethylene materials. These materials are known to be highly resistant to natural degradation.

3.3.1. SEM analysis

Scanning Electron Microscopy (SEM) provided crucial insights into the surface alterations of low-density polyethylene (LDPE) films. The untreated control samples exhibited smooth, homogeneous surfaces with no evidence of deterioration, highlighting the inherent stability and resistance of LDPE to environmental conditions (Fig. 6A). In contrast, enzyme-treated films display substantial surface etching, debris shedding, and the formation of structural anomalies, such as ridges, depressions, holes, and pores. These changes are indicative of enzymatic activity that breaks down the polymer matrix. The visible fractures and uneven surfaces were consistent with previous observations, such as the degradation of LDPE by a mixed culture of *Lysinibacillus xylanilyticus* and *Aspergillus niger*, as well as HDPE degradation by *Cephalosporium* species. These results strongly suggest that the applied enzyme mixtures were effective in attacking and degrading the LDPE polymer chains.

3.3.2. ATR-FTIR spectroscopy

ATR-FTIR spectroscopy was used to investigate the chemical changes in the LDPE films. The enzyme-treated samples showed significant alterations in their infrared spectra compared those with of the control (Fig. 6B). Untreated LDPE typically exhibits distinct stretching and bending vibrations of the -CH2 groups in the primary IR band, particularly at 2954 cm^{-1} and 2843 cm^{-1} , characteristic of its ethylene backbone. The treated samples demonstrated reduced intensity of these peaks, indicating a loss of hydrophobic chains due to enzymatic action.

Additionally, new carbonyl functional groups emerged within the range of 1600–1850 cm^{-1} in the enzyme-treated films. This carbonyl band is indicative of oxidative degradation, suggesting that the enzymes facilitate the breakdown of long polymer chains into smaller fragments containing oxygenated groups. These findings align with those reported by Khruengsai et al. (2021);Khruengsai et al. (2021), and further confirm the structural transformations caused by enzymatic action. The increased peak intensities and emergence of new functional groups reinforce the conclusion that the enzyme mixture effectively catalyzed the degradation of LDPE.

3.3.3. AFM analysis

Atomic Force Microscopy (AFM) provided detailed topographical information on the treated and untreated LDPE surfaces. The control samples exhibited minimal variation in surface height, with a maximum roughness of 205 nm, reflecting the structural integrity of untreated LDPE (Fig. 6C). In contrast, the enzyme-treated films displayed dramatic changes with the formation of grooves, pits, cracks, and significant erosion. The maximum surface height variation increased to 407 nm, underscoring the effectiveness of enzymatic action in altering the polymer surface (Rani et al., 2022).

The observed topographical changes strongly indicate that the enzymes selectively targeted the LDPE matrix, breaking down polymer chains and creating localized areas of degradation. Similar surface morphologies have been reported in studies involving LDPE degradation by *Bacillus cereus* and polyvinyl chloride degradation by bacterial consortia isolated from *Tenebrio molitor* larvae. This consistency across studies underscores the potential of enzymatic treatment for degrading resilient plastic polymers.

3.3.4. Comparative insights and practical implications

The combined findings from SEM, ATR-FTIR, and AFM analyses conclusively demonstrated the capability of plastic-degrading enzymes to effectively degrade LDPE under controlled laboratory conditions.

While these laboratory-based findings are highly promising, it is important to acknowledge that environmental factors such as soil moisture, nutrient availability, temperature, and pollutant loads significantly influence enzymatic activity. In natural ecosystems, synergistic interactions between diverse microbial communities and their enzymes play a critical role in plastic degradation. However, laboratory experiments, constrained by the limitations of culturing methods, can isolate only a few microorganisms present in natural environments. Despite these limitations, our results demonstrated that even enzymes derived from a limited microbial subset can exhibit significant LDPE degradation under controlled conditions. These findings provide a solid foundation for further exploration under simulated and real-world environmental conditions to optimize enzyme activity and assess their practical bioremediation potential (Nowak et al., 2011).

4. Conclusion

This study significantly contributes to the advancement in the field of environmental catalysis by identifying novel PET-degrading enzymes from soil samples heavily contaminated with plastics. Utilizing a combination of cutting-edge computational models such as Generative Adversarial Networks for data augmentation and Evolutionary Scale Modeling for sequence classification, we expanded the known repertoire of enzymes capable of plastic degradation. These enzymes represent a diverse array of previously undiscovered plastizymes, underscoring the potential of metagenomic approaches for finding solutions to environmental challenges. Our findings not only contribute to the understanding of enzymatic plastic degradation but also present a novel methodology for identifying and characterizing potential biocatalysts. This approach is crucial for the development of more efficient and environment-friendly techniques for managing plastic waste. The discovery of these enzymes is a step towards innovative biotechnological applications in plastic waste management. Our research suggests the need for further exploration of the scalability of these biodegradation processes, including their economic viability and environmental impact. The long-term goal is to integrate these novel enzymes into practical applications to mitigate plastic pollution, thus contributing to a more sustainable and cleaner environment.

CRediT authorship contribution statement

Mohammad Reza Rezaei Barzan: Writing – original draft, Validation, Software, Resources, Formal analysis. **Donya Afshar Jahanshahi:** Writing – original draft, Visualization, Validation, Methodology, Investigation. **Kaveh Kavousi:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Shohreh Ariaeenejad:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization. **Mohammad Bahram:** Writing – review & editing, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was carried out by the support of the Institute of Biochemistry and Biophysics (IBB), University of Tehran and Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran (ABRII).

Code availability

All code for analysis associated with the current study is available at https://github.com/mohrezaeib/plastizyme-miner.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ecoenv.2024.117640.

Data Availability

Data will be made available on request.

References

Ariaeenejad, S., Kavousi, K., Afshar Jahanshahi, D., Sheykh Abdollahzadeh Mamaghani, A., Ghasemitabesh, R., Moosavi-Movahedi, A.A., Hosseini Salekdeh, G., 2022a. Enzymatically triggered delignification through a novel stable laccase: a mixed in-silico /in-vitro exploration of a complex environmental microbiota. Int. J. Biol. Macromol. 211, 328–341. https://doi.org/10.1016/j.ijbiomac.2022.05.039.

Ariaeenejad, S., Kavousi, K., Han, J-L., Ding, X-Z, Salekdeh, G.H., 2023. Efficiency of an alkaline, thermostable, detergent compatible, and organic solvent tolerant lipase with hydrolytic potential in biotreatment of wastewater. Sci. Total Environ.

Ariaeenejad, S., Kavousi, K., Mamaghani, A.S.A., Ghasemitabesh, R., Hosseini Salekdeh, G., 2022b. Simultaneous hydrolysis of various protein-rich industrial wastes by a naturally evolved protease from tannery wastewater microbiota. Sci. Total Environ. 815, 152796. https://doi.org/10.1016/j.scitotenv.2021.152796.

Ariaeenejad, S., Kavousi, K., Zeinalabedini, M., Afshar Jahanshahi, D., Beh-Afarin, S.R., 2024. Enhanced bioremediation of saline azo dye effluents using PersiLac3: a thermo-halotolerant laccase from a tannery metagenome. Results Eng. 24, 103172. https://doi.org/10.1016/j.rineng.2024.103172.

Austin, Harry Patrick, Allen, M.D., Donohoe, B.S., Rorrer, N.A., Austin, Harry P., Allen, M.D., Donohoe, B.S., Rorrer, N.A., Kearns, F.L., Silveira, R.L., Pollard, B.C., Dominick, G., Duman, R., El, K., Mykhaylyk, V., Wagner, A., 2018. Characterization and engineering of a plastic-degrading aromatic polyesterase. https://doi.org/10.1073/pnas.1718804115.

Charnock, C., 2021. Norwegian soils and waters contain mesophilic, plastic-degrading bacteria. Microorganisms 9, 1–18. https://doi.org/10.3390/microorganisms9010094.

Ekanayaka, A.H., Tibpromma, S., Dai, D., Xu, R., Suwannarach, N., Stephenson, S.L., Dao, C., Karunarathna, S.C., 2022. A review of the fungi that degrade plastic. J. Fungi 8, 1–27. https://doi.org/10.3390/jof8080772.

Foroozandeh Shahraki, M., Ariaeenejad, S., Fallah Atanaki, F., Zolfaghari, B., Koshiba, T., Kavousi, K., Salekdeh, G.H., 2020. MCIC: automated identification of cellulases from metagenomic data and characterization based on temperature and pH dependence. Front. Microbiol. 11, 1–10. https://doi.org/10.3389/fmicb.2020.567863.

Foroozandeh Shahraki, M., Farhadyar, K., Kavousi, K., Azarabad, M.H., Boroomand, A., Ariaeenejad, S., Hosseini Salekdeh, G., 2021. A generalized machine-learning aided method for targeted identification of industrial enzymes from metagenome: a xylanase temperature dependence case study. Biotechnol. Bioeng. 118, 759–769. https://doi.org/10.1002/bit.27608.

Gupta, A., Agrawal, S., 2023. Machine learning-based enzyme engineering of PETase for improved efficiency in plastic degradation. J. Emerg. Investig. 6, 1–10. https://doi.org/10.59720/22-016.

Hashim Khudhair, Z., Hosseini, M., Ghanbari, M., Mahdi, M.A., Jasim, L.S., Salavati-Niasari, M., 2023. Sonochemical synthesis and characterization of Mn2+Mn3+6 [SiO4]O8 nanostructures and their application as photocatalyst for degradation of water-soluble organic pollutants below simulated sunlight. Arab. J. Chem. 16, 105132. https://doi.org/10.1016/j.arabjc.2023.105132.

Jahanshahi, D.A., Ariaeenejad, S., Kavousi, K., 2023. A metagenomic catalog for exploring the plastizymes landscape covering taxa, genes, and proteins. Sci. Rep. 13, 1–14. https://doi.org/10.1038/s41598-023-43042-9.

Joo, S., Cho, I.J., Seo, H., Son, H.F., Sagong, H., Shin, T.J., Choi, S.Y., Lee, S.Y., Kim, K., n.d. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. Nat. Commun. https://doi.org/10.1038/s41467-018-02881-1.

Kawai, F., Kawabata, T., Oda, M., 2019. Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. Appl. Microbiol. Biotechnol. 103, 4253–4268. https://doi.org/10.1007/s00253-019-09717-y.

Khruengsai, S., Sripahco, T., Pripdeevech, P., 2021. Low-Density Polyethylene Film Biodegradation Potential by Fungal Species from Thailand.

Kitadokoro, K., Kakara, M., Matsui, S., Osokoshi, R., Thumarat, U., 2019. Structural insights into the unique polylactate-degrading mechanism of Thermobifida alba cutinase, 286, 2087–2098. https://doi.org/10.1111/febs.14781.

Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., Zhou, X., 2020. Generative adversarial networks and its applications in biomedical informatics. Front. Public Heal. 8, 1–14. https://doi.org/10.3389/fpubh.2020.00164.

Lear, G., Kingsbury, J.M., Franchini, S., Gambarini, V., Maday, S.D.M., Wallbank, J.A., Weaver, L., Pantos, O., 2021. Plastics and the microbiome: impacts and solutions. Environ. Micro 16, 1–19. https://doi.org/10.1186/s40793-020-00371-w.

Li, Z., Wei, R., Gao, M., Ren, Y., Yu, B., Nie, K., Xu, H., Liu, L., 2020. Biodegradation of low-density polyethylene by Microbulbifer hydrolyticus IRE-31. J. Environ. Manag. 263. https://doi.org/10.1016/j.jenvman.2020.110402.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., Rives, A., 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379 (80), 1123–1130. https://doi.org/10.1126/science.ade2574.

Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H., Luo, Y., 2018. Protein Crystallography and Site-Direct Mutagenesis Analysis of the Poly(ethylene terephthalate) Hydrolase PETase from Ideonella sakaiensis. https://doi.org/10.1002/cbic.201800097.

Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., Wang, Z., 2019. Wasserstein GAN-Based Small-Sample Augmentation for New-Generation Artificial Intelligence: A Case Study of Cancer-Staging Data in Biology. Engineering 5, 156–163. https://doi.org/10.1016/j.eng.2018.11.018.

Mican, J., Jaradat, D.M.M., Liu, W., Weber, G., Mazurenko, S., Bornscheuer, U.T., Damborsky, J., Wei, R., Bednar, D., 2024. Exploring new galaxies: Perspectives on the discovery of novel PET-degrading enzymes. Appl. Catal. B Environ. 342, 123404. https://doi.org/10.1016/j.apcatb.2023.123404.

Motamedi, E., Kavousi, K., Sadeghian Motahar, S.F., Ghaffari, M.R., Sheykh Abdollahzadeh Mamaghani, A., Hosseini Salekdeh, G., Ariaeenejad, S., 2021. Efficient removal of various textile dyes from wastewater by novel thermo-halotolerant laccase. Bioresour. Technol., 125468 https://doi.org/10.1016/j.biortech.2021.125468.

Nowak, B., Pajak, J., Drozd-Bratkowicz, M., Rymarz, G., 2011. Microorganisms participating in the biodegradation of modified polyethylene films in different soils under laboratory conditions. Int. Biodeterior. Biodegrad. 65, 757–767. https://doi.org/10.1016/j.ibiod.2011.04.007.

Nutaratat, P., Arigul, T., Srisuk, N., Kruasuwan, W., 2024. Microbiome sequencing revealed the abundance of uncultured bacteria in the Phatthalung sago palm-growing soil. PLoS One 19, 1–12. https://doi.org/10.1371/journal.pone.0299251.

Otzen, M., Palacio, C., Janssen, D.B., 2018. Characterization of the caprolactam degradation pathway in Pseudomonas jessenii using mass spectrometry-based proteomics. Appl. Microbiol. Biotechnol. 102, 6699–6711. https://doi.org/10.1007/s00253-018-9073-7.

Palm, G.J., Reisky, L., Böttcher, D., Müller, H., Michels, E.A.P., Walczak, M.C., Berndt, L., Weiss, M.S., Bornscheuer, U.T., Weber, G., 2019. Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. Nat. Commun. 10, 1–10. https://doi.org/10.1038/s41467-019-09326-3.

Pathak, V.M., Navneet, 2017. Review on the current status of polymer degradation: a microbial approach. Bioresour. Bioprocess. 4. https://doi.org/10.1186/s40643-017-0145-9.

Prijambada, I.D., Negoro, S., Yomo, T., Urabe, I., 1995. Emergence of nylon oligomer degradation enzymes in Pseudomonas aeruginosa PAO through experimental evolution. Appl. Environ. Microbiol. 61, 2020–2022. https://doi.org/10.1128/aem.61.5.2020-2022.1995.

Rani, R., Rathee, J., Kumari, P., Singh, N.P., Santal, A.R., 2022. Biodegradation and detoxification of low-density polyethylene by an indigenous strain Bacillus licheniformis SARR1. J. Appl. Biol. Biotechnol. 10, 9–21. https://doi.org/10.7324/JABB.2021.100102.

Réjasse, Agnès, Waeytens, Jehan, Deniset-Besseau, Ariane, Crapart, Nicolas, Nielsen-Leroux, Christina, 2022. and C.S., n.d. plastic biodegradation: do Galleria mellonella larvae bioassimilate polyethylene? A spectral histology approach using isotopic labeling and infrared microspectroscopy. Environ. Sci. Technol. 56 (1), 525–534. https://doi.org/10.1021/acs.est.1c03417.

Repecka, D., Jauniskis, V., Karpus, L., Rembeza, E., Rokaitis, I., Zrimec, J., Poviloniene, S., Laurynenas, A., Viknander, S., Abuajwa, W., Savolainen, O., Meskys, R., Engqvist, M.K.M., Zelezniak, A., 2021. Expanding functional protein sequence spaces using generative adversarial networks. Nat. Mach. Intell. 3, 324–333. https://doi.org/10.1038/s42256-021-00310-5.

Rodriguez-Conde, S., Molina, L., González, P., García-Puente, A., Segura, A., 2016. Degradation of phenanthrene by Novosphingobium sp. HS2a improved plant growth in PAHs-contaminated environments. Appl. Microbiol. Biotechnol. 100, 10627–10636. https://doi.org/10.1007/s00253-016-7892-y.

Samak, N.A., Jia, Y., Sharshar, M.M., Mu, T., Yang, M., Peh, S., Xing, J., 2020. Recent advances in biocatalysts engineering for polyethylene terephthalate plastic waste green recycling. Environ. Int. 145, 106144. https://doi.org/10.1016/j.envint.2020.106144.

Shahraki, Mehdi F., F.F.A.S.A.M.R.G.M.H.N.-B.M.M.G.H.S.K.K, 2022. A computational learning paradigm to targeted discovery of biocatalysts from metagenomic data: a case study of lipase identification. Biotechnol. Bioeng. 119.

Sowmya, H.V., Ramalingappa, Krishnappa, M., Thippeswamy, B., 2015. Degradation of polyethylene by Penicillium simplicissimum isolated from local dumpsite of Shivamogga district. Environ. Dev. Sustain. 17, 731–745. https://doi.org/10.1007/s10668-014-9571-4.

Tao, X., Ouyang, H., Zhou, A., Wang, D., Matlock, H., Morgan, J.S., Ren, A.T., Mu, D., Pan, C., Zhu, X., Han, A., Zhou, J., 2023. Polyethylene degradation by a rhodococcous strain isolated from naturally weathered plastic waste enrichment. Environ. Sci. Technol. 57, 13901–13911. https://doi.org/10.1021/acs.est.3c03778.

Wu, B., Zhong, B., Zheng, L., Huang, R., Jiang, S., 2024. ϵ Harnessing Protein Language Model for Structure-Based Discovery of Highly Efficient and Robust PET Hydrolases.

Yousefzadeh, F., Ghanbari, M., Salavati-Niasari, M., 2023. Sonochemical synthesis and characterization of Sm2CuO4 nanostructures and their application as visible-light photocatalyst for degradation of water-soluble organic pollutants. Chemosphere 338, 139564. https://doi.org/10.1016/J.CHEMOSPHERE.2023.139564.