Using Metagenomic Analysis to uncover the Phylogeny of Xenobiotic Degrading Compounds found near the San Jacinto River

By: Adib Shafipour

Abstract

The prevalence of recalcitrant compounds, compounds which are slow to biodegrade, is becoming a growing concern as it causes harm to the ecology. The ability of xenobiotic degrading compounds cannot sufficiently breakdown the newly introduced recalcitrant compounds at the current rate. Polyethylene plastics are an example of some of the recalcitrant compounds which are commonly found in rivers and soil. Polyethylene is widely used, 'over 100 million tons of polyethylene resins are produced annually making up a total of 34% of all plastics in use' (Farber). The process of breaking down recalcitrant compounds such as Polyethylene plastics can be gleaned by different processes in metagenomics. Metagenomics uses samples collected directly from the environment to understand the composition of existing xenobiotic degrading compounds by using bioinformatics tools to aid in our construction of the phylogeny. Samples were gathered from 3 separate locations which run along the San Jacinto River with similar soil construction. The San Jacinto River was widely used as a waste dump for many xenobiotic compounds including many polyethylene plastics. Each of the samples were retrieved on soil rich sediments from the river. Using this analysis, we will be able to find enzymes from three distinct set of bacterial organisms which are able to degrade polyethylene via bioinformatics tools. The likelihood that these samples have bacteria which can degrade polyethylene plastics is high because of the amount of polyethylene deposits that are placed in different locations along the river. These discovered enzymes give insight into the process by which polyethylene plastic degradation occurs in local marine environments.

Introduction

Six bacterial samples were taken from three separate locations on the San Jacinto River, near central Houston, which had their metagenome analyzed via bioinformatics software tools in order to understand more about the nature of their composition. This analysis is to detect if these bacteria have genes that enable them to degrade xenobiotic compound, if so, sequential and phylogenetic analysis will be performed on the relevant genes. The samples are named as follows, WGM-01_R1_001 ,WGM-01_R2_001, WGM-04_R1_001, WGM-04_R2_001, WGM-07_R1_001, WGM-07_R2_001 relative to their location in order to distinguish specific characteristics based on location.

Figure 1: Location of Sample Collection Data in Houston

These tools can be used to apply the xenobiotic degrading bacteria in order to locate specific proteins that can be isolated and used to degrade these xenobiotic compounds, specifically polyethylene found in plastics. We will conduct RAST analysis using two distinct softwares in order to annotate the genome of the sample. These processes will test if the 3 distinct locations selected along the San Jacinto River are going to give distinct phylogenetic data that distinguishes the proteins necessary for the degradation of polyethylene, specifically the type of proteins and pathways used in the breakdown of these compounds. The xenobiotic compounds deposited in close proximity to the sediments contain the degradation proteins of the polyethylene plastics that could be isolated. The analysis should show 3 distinct bacterial compositions with a common xenobiotic compound degrading enzyme.

Materials and Methods

The two main bioinformatics software used to conduct this analysis were MG-RAST and BLAST2go. These are common open source resources which allow researchers to use global database in order to help construct phylogeny and metagenomic makeup of the sample organism retrieved from the San Jacinto River. After specific organisms are identified BLAST will be performed in order to identify any proteins which are present within specific segments of the sample.

MG-RAST

One of the most widely used bioinformatics tools, RAST can be used to annotate samples from the environment using various processes. “MG-RAST, or the metagenomics RAST (rapid annotations using subsystems technology) server makes it possible for users to upload raw

metagenomic sequence data in (preferably) fastq or fasta format" (Keegen). This program stores and process metagenomic data by comparing sequence via nucleotide arrangement or amino acid arrangement. After this arrangement is performed, the program can archive the data, annotate the sequences, and analyze the function of the metagenomic data.

BLAST2Go

Blast2go is a basic local alignment search tool that allows us to annotate the various samples. BLAST2GO is another program used to annotate the sample. It is more visual descriptive than MG-RAST , as it can show graphically the relationships between the various phyla found on each of the samples.This program however is not free to use and is relatively more difficult to navigate in terms of obtaining specific information on hits.

NCBI BLAST tool

The National Institute of Biotechnology Information (NCBI) is a database of biotechnology information on the protein and nucleotide composition of organisms that contains tools for bioinformatics analysis. The Basic Local Alignment search tool (BLAST) is a tool used to analyze nucleotide or protein sequences to categorize them based on uploaded information contained in the database that can describe their composition and function.

Results and Discussion

After an analysis of each of the six sequences were completed, The data showed a sequential breakdown of the known proteins found in the database. The average of each of the six sequences shows that 65.00% of the proteins found in the sample are unknown, by not matching any sequence in the database which is rather unusual because it cannot be traced to any type of sequence within the NCBI database. This is to be expected for all of the other sediments because these proteins are relatively unannotated on the database. Further analysis into the specific nature of all of the found proteins would be required in order to find specific functional qualities of the unknown sequences. The data also showed that there was only 0.06% of ribosomal RNA, indicating that there is a low amount of protein synthesis occurring in the samples in their natural environment. This would suggest that the bacteria contain a higher degree of metabolic activity because it does not have a high activity of protein synthesis. Each of the samples contained over 8 billion base pairs with about 54 million sequences per sample.

The protein features that are present within the samples exceeds 47 million hits, which encode various data that is necessary for the bacteria to degrade the prevalent xenobiotic compounds. The average of these samples as indicated by BLAST2Go in Figure 2, shows us the mean GC percentage between the six samples was near 64% with a standard deviation of 8% relative to each of the six samples. This means that these samples have a much higher annealing temperature when Polymerase Chain Reaction is performed in order to make copies of the proteins found in these samples.

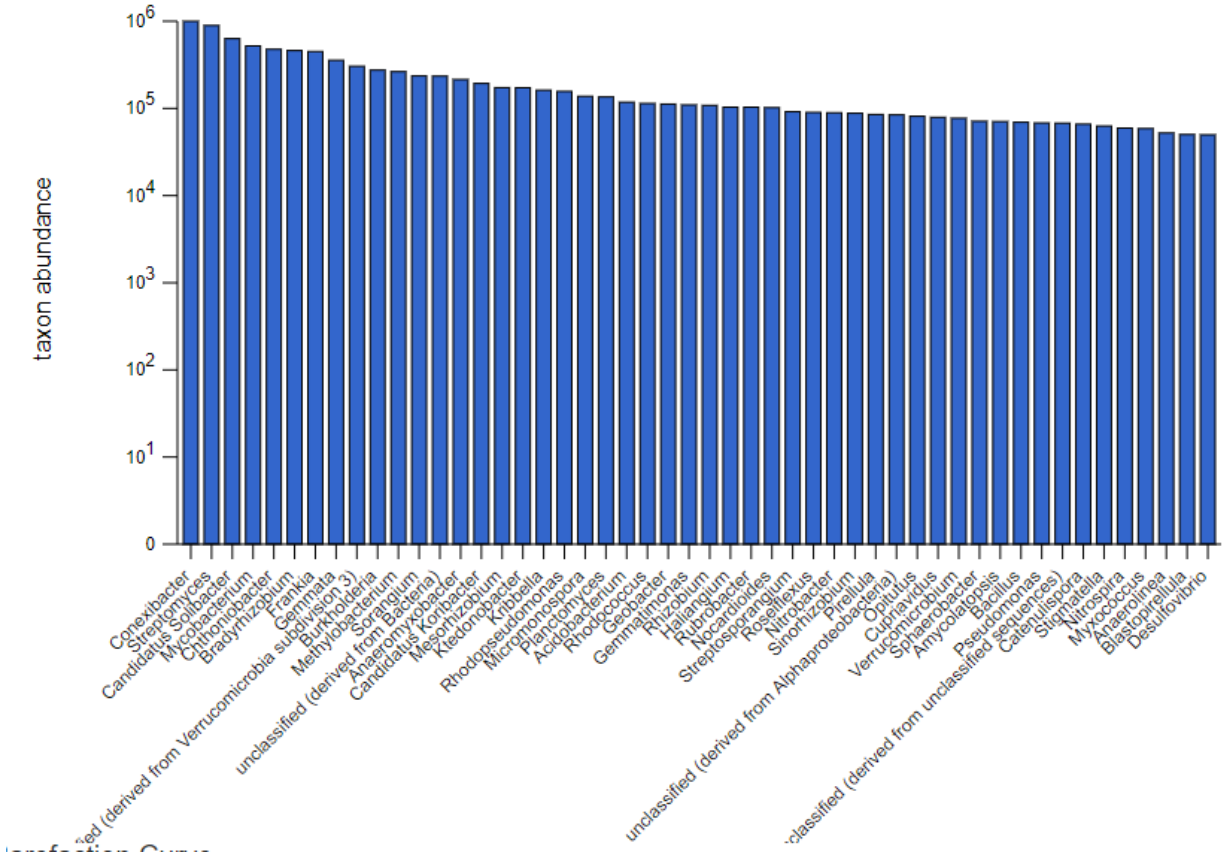


Figure 2: Average of 6 Samples Taxon Abundance from BLAST2Go

The various phyla found in the sample ranges from Conexibacter to Plantomyces this would suggest that the phyla were not phylogenetically linked as they are not closely related evolutionarily. The abundance in the sample suggests a few things about the nature of the sample, as it contains over 16,000 species within its genomic data. This would suggest that these different species play different roles in the functioning of these samples that are not all related to the xenobiotic degradation of polyethylene. Further isolation of specific species present in the sample would indicate more clearly the function of each of the species in the xenobiotic degradation processes, as well as further potential for pathway discoveries that allow this process to occur.

The data in this sample shows the makeup of the protein because less of the functions of the proteins are unclear in that the number of reads of the sample by RAST has shown that over 16,000 species of bacteria have been identified. This indicates that the samples found at the sediments of the river contain various

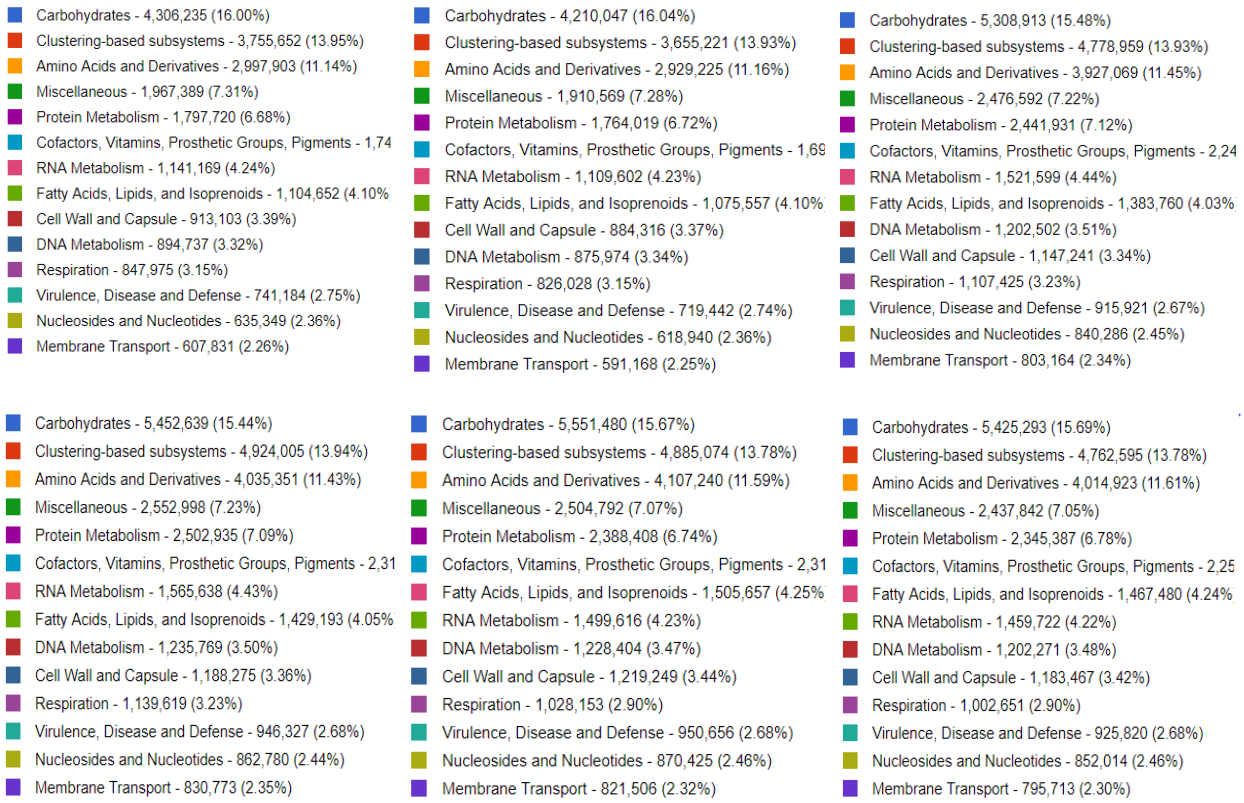
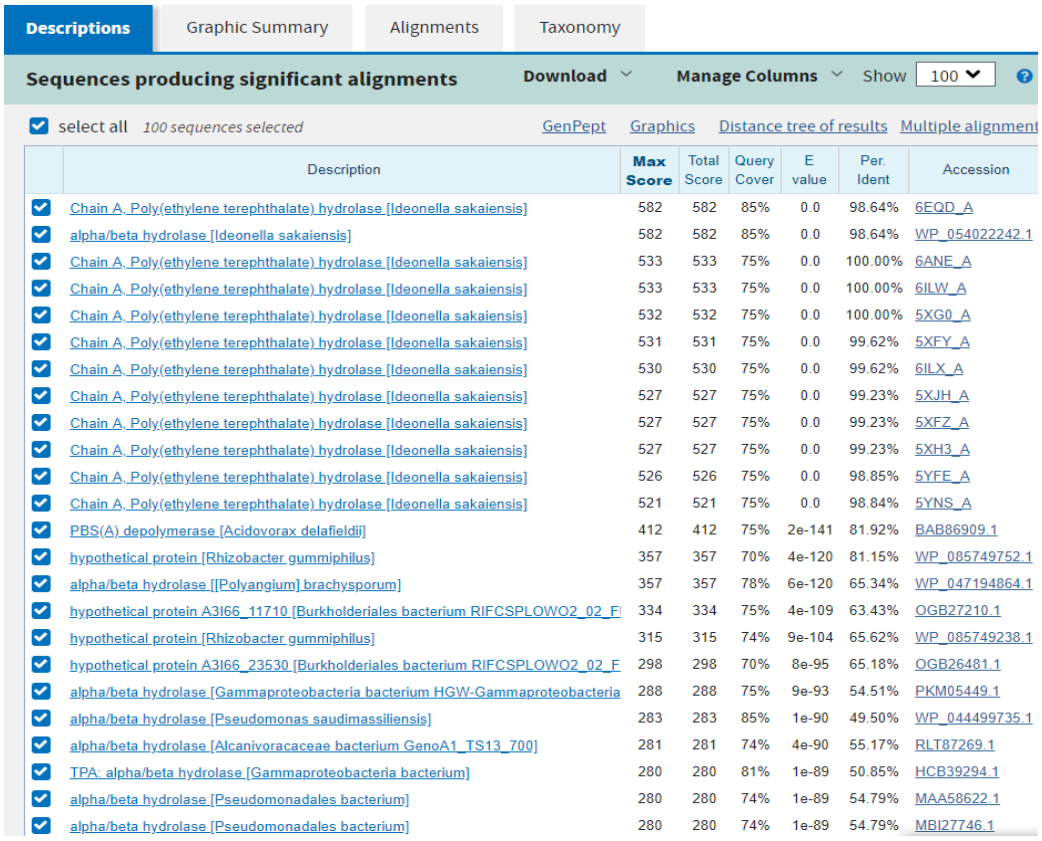


Figure 3: 6 Sample Subsystems Layout Percentages (from upper left to bottom right) : WGM-01_R1_001 ,WGM-01_R2_001, WGM-04_R1_001, WGM-04_R2_001, WGM-07_R1_001, WGM-07_R2_001

This figure shows the various subsystems which are contained in the samples. This allotment of subsystems seems to be universal throughout the given samples with a standard deviation of the subsystems being of less than 1% from each other. This shows that although the six samples were derived from three separate locations albeit along the same river system, that the genes necessary to degrade xenobiotic compounds, and more specifically for the polyethylene plastics do not vary to a large extent.

The diversity of the phyla that were detected in the analysis. The sequence seems to be primarily within the phylum of proteobacteria in which 45.91% and actinobacteria which is 29.30% of the organism. Proteobacteria is known for being gram-negative in that it exhibits characteristics of a cell wall which contains peptidoglycan. While actinobacteria are gram positive bacteria which do not contain the cell wall of peptidoglycan. Due to the more active nature of proteobacteria being more likely to contain hydrolysis performing enzymes, analysis focused on the phyla of proteobacteria within the subsystem of protein metabolism in order to clearly identify enzymes which can cause the hydrolysis of polyethylene leading to their degradation.

Initial protein metabolism analysis was performed on the WGM-04_R1_001 sample because it had the highest ratio and overall amount of protein metabolism at just over 2.4 million base pairs. The contents of the protein metabolism subsystem ran through BLAST on the NCBI open source database to sequence the enzymes present within the sample.



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Manage Columns — Show 100

select all 100 sequences selected — GenPept — Graphics — Distance tree of results — Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	582	582	85%	0.0	98.64%	6EQD_A
alpha/beta hydrolase [Ideonella sakaiensis]	582	582	85%	0.0	98.64%	WP_054022242.1
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	533	533	75%	0.0	100.00%	6ANE_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	533	533	75%	0.0	100.00%	6ILW_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	532	532	75%	0.0	100.00%	5XG0_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	531	531	75%	0.0	99.62%	5XFY_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	530	530	75%	0.0	99.62%	6ILX_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	527	527	75%	0.0	99.23%	5XJH_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	527	527	75%	0.0	99.23%	5XFZ_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	527	527	75%	0.0	99.23%	5XH3_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	526	526	75%	0.0	98.85%	5YFE_A
Chain A, Poly(ethylene terephthalate) hydrolase [Ideonella sakaiensis]	521	521	75%	0.0	98.84%	5YNS_A
PBS(A) depolymerase [Acidovorax delafieldii]	412	412	75%	2e-141	81.92%	BAB86909.1
hypothetical protein [Rhizobacter gummiphilus]	357	357	70%	4e-120	81.15%	WP_085749752.1
alpha/beta hydrolase [[Polyangium] brachysporum]	357	357	78%	6e-120	65.34%	WP_047194864.1
hypothetical protein A3I66_11710 [Burkholderiales bacterium RIFCSPLOWO2_02_F	334	334	75%	4e-109	63.43%	OGB27210.1
hypothetical protein [Rhizobacter gummiphilus]	315	315	74%	9e-104	65.62%	WP_085749238.1
hypothetical protein A3I66_23530 [Burkholderiales bacterium RIFCSPLOWO2_02_F	298	298	70%	8e-95	65.18%	OGB26481.1
alpha/beta hydrolase [Gammaproteobacteria bacterium HGW-Gammaproteobacteria	288	288	75%	9e-93	54.51%	PKM05449.1
alpha/beta hydrolase [Pseudomonas saudimassiliensis]	283	283	85%	1e-90	49.50%	WP_044499735.1
alpha/beta hydrolase [Alcanivoracaceae bacterium GenoA1_TS13_700]	281	281	74%	4e-90	55.17%	RLT87269.1
TPA: alpha/beta hydrolase [Gammaproteobacteria bacterium]	280	280	81%	1e-89	50.85%	HCB39294.1
alpha/beta hydrolase [Pseudomonadales bacterium]	280	280	74%	1e-89	54.79%	MAA58622.1
alpha/beta hydrolase [Pseudomonadales bacterium]	280	280	74%	1e-89	54.79%	MBI27746.1

Figure 4: NCBI BLAST data on Protein Metabolism of Sample WGM-04_R1_001

The isolation of the sample showed that Ideonella sakaiensis was the most easily identifiable bacterial organism that was found in the protein metabolism subsystem. Ideonella sakaiensis is an aerobic, rod shaped bacteria that has been found to degrade plastics from recent studies. Specific strains have been found to be “capable of hydrolyzing PET and the reaction intermediate, mono(2-hydroxyethyl) terephthalic acid.” (Yoshida). Analysis of the 5 other samples specifically with intent to isolate any traces of Ideonella sakaiensis, each of the samples contained a near same amount of the Ideonella sakaiensis strain relative to the sample size. This indicates the strains have a near similar genetic composition that includes polyethylene degradation enzymes.

Deeper analysis of the specific protein function of the enzymes of the species revealed that the strain of I. sakaiensis contained Polyethylene terephthalate hydrolase (PETase). Bacterial organism hydrolases signify the catalization of an enzyme which causes hydrolysis of Polyethylene in this case. Research on PETase has recently discovered due to the recent discovery of the specific nature of the PET degrading organism which was “found near a recycled plastic bottle factory in Japan’ (Yoshida).

RecName: Full=Poly(ethylene terephthalate) hydrolase; Short=PET hydrolase; Short=PETase; AltName: Full=PET-digesting enzyme; Flags: Precursor

UniProtKB/Swiss-Prot: A0A0K8P6T7.1

GenPept Identical Proteins Graphics

>sp|A0A0K8P6T7.1|PETH_IDESA RecName: Full=Poly(ethylene terephthalate) hydrolase; Short=PET hydrolase; Short=PETase; AltName: Full=PET-digesting enzyme; Flags: Precursor
MNFPRASRLMQAAVLGGLMAVSAAATAQTNPYARGPNPTAASLEASAGPFTVRSFTVSRPSGYGAGTVYY
PTNAGGTVGAIAIVPGYTARQSSIKWWGPRLASHGFVVITIDTNSTLDQPSSRSSQQMAALRQVASLNGT
SSSPIYGKVDTARMGVMGWSMGGGGSLISAANNPSLKAAAPQAPWDSSTNFSSVTVPTLIFACENDSIAP
VNSSALPIYDSMSRNAKQFLEINGGSHSCANSGNSNQALIGKKGVAWMKRFMDNDTRYSTFACENPNSTR
VSDFRTANCS

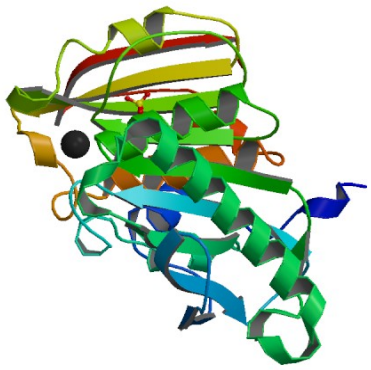


Figure 5: Identification of amino acid composition of Polyethylene terephthalate hydrolase with image graphic

The sequence above was isolated in each of the six samples, indicating that the samples contained the PET digesting enzyme which was actively used in the degradation of the plastics in the sludge samples. Further understanding of the process of the PETase to degrade polyethylene shows that the PETase will cleave the plastic into monohydroxyethyl terephthalate hydrolase (MHETase) and other terephthalic acids which are typically used to engineer polymers. These terephthalic acids are neutral to the environment, which cause no harm to the marine ecology. This reaction system is undergone at a slow rate causing the long time degradation of the existing PET compounds.

Conclusion

In summary, a phylogenetic analysis of the six samples showed that although they contain an abundance of different species that are used to degrade xenobiotic compounds, they do not vary greatly in terms of their composition even though they were located in various sediment outputs along the San Jacinto River. The samples were not geographically distinct and showed no signs of deep phylogeny within the species that are contained in it. All samples exhibited similar composition , and varied minimally in their genetic properties. The specific enzyme present in each of the six samples that degrade polyethylene was PETase which was found to be nearly consistent ratios of the total sample sequences within 0.1%. The samples contained almost completely similar subsystem and bacterial strain composition, contrary to the initial premise that the three distinct locations would contain samples with differing bacterial compositions. All samples contained a similar polyethylene degrading enzyme PETase which is a recently isolated hydrolase of PET.

Applications

Some practical applications could before industry to use these xenobiotic degrading compounds to minimize impact on the ecology should waste be a byproduct of their industrial developmental processes. An effort to make these polyethylene plastics more, “biodegradable by developing microorganisms capable of degrading the compound and by treating the compound to

make it more conducive to microbial attack." (Geyer). Further research on the nature of multiplying the PETase with specific intentions of using the enzyme to degrade plastics in the environment for sustained environmentally neutral outcomes would be required. This enzyme could make biodegradation via bacterial samples a means of recycling plastics reducing costs and improving environmental health

Citations

Geyer, Roland; Jambeck, Jenna R.; Law, Kara Lavender (1 July 2017). "Production, use, and fate of all plastics ever made". Science Advances. 3 (7): e1700782. Bibcode:2017SciA....3E0782G. doi:10.1126/sciadv.1700782. PMC 5517107. PMID 28776036.

Farber, Marcel. "Microbial Degradation of Recalcitrant Compounds and Synthetic Aromatic Polymers." Enzyme and Microbial Technology, Elsevier, 13 Nov. 2002, www.sciencedirect.com/science/article/pii/0141022979900413.

Keegan, Kevin P, et al. "MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function." Methods in Molecular Biology (Clifton, N.J.), U.S. National Library of Medicine, 2016, www.ncbi.nlm.nih.gov/pubmed/26791506.

Iyer R, Damania A, Iken B. Genome data of Stenotrophomonas maltophilia DF07 collected from polluted river sediment reveals an opportunistic pathogen and a potential antibiotic reservoir. Data Brief. 2019;25:104137. Published 2019 Jun 11. doi:10.1016/j.dib.2019.104137

Yoshida S, Hiraga K, Takehana T, Taniguchi I, Yamaji H, Maeda Y, et al. (March 2016). "A bacterium that degrades and assimilates poly(ethylene terephthalate)". Science. 351 (6278): 1196–9. doi:10.1126/science.aad6359