



Shotgun metagenomic analysis reveals the diversity of PHA producer bacterial community and PHA synthase gene in Addis Ababa municipal solid waste disposal area 'Qoshe'

Zuriash Mamo^{1,2}, Sewunet Abera^{3,4,5} and Mesfin Tafesse^{1,2*} 

Abstract

Background Polyhydroxyalkanoates (PHAs) are naturally produced biopolymers with significant scientific and biotechnological potential. This study aimed to investigate the diversity of the PHA-producing bacterial community and PhaC genes in soil samples collected from a municipal solid waste disposal site known as "Qoshe" in Addis Ababa, Ethiopia, using a shotgun metagenomics approach. The SqueezeMeta pipeline was used to analyze the microbial community in the waste samples. A CD search against the TIGRFAM protein family database was performed to identify the complete-length multidomain sequences of PhaC genes and classify them into their respective classes. Statistical analysis and data visualization were performed using RStudio with R version 4.2.3.

Results The findings of this study suggest that known and unknown taxa likely contribute to the phaC genes of municipal solid waste. Taxonomic profiling of the metagenomic data revealed that the majority of the PHA-producing taxa belonged to the phylum Proteobacteria (80%), followed by Actinomycetota (16.5%). Furthermore, this study identified *Thiomonas* and unclassified *Mycobacterium* as the main contributors to class I PhaC genes. Class II PhaC genes are predominantly associated with the *Pseudomonadaceae* family, followed by unclassified *Hyphomicrobiales* and *Acidimicrobiales*. Class III PhaC genes are abundantly related to the *Methylococcaceae* family, specifically the *Methylocaldum* genus. The analysis of PhaC gene sequences revealed high level of diversity, with a significant proportion of putative PhaC genes exhibiting low sequence identity with each other and PhaC gene in the database. Notably, the sequence variation observed within the same PhaC gene classes suggests the potential presence of previously unidentified PhaC gene variants.

Conclusions Overall, this research improves our understanding of the diversity of PHA-producing taxa and PhaC genes in municipal solid waste environments, providing opportunities for sustainable PHA production and waste

*Correspondence:

Mesfin Tafesse
mesfin.tafesse@aastu.edu.et

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

management strategies. However, additional studies, including the isolation and characterization of specific strains, are necessary to confirm the PHA production capabilities of these strains and explore their biotechnological potential.

Keywords Polyhydroxyalkanoates, Municipal solid waste, Bacterial community, Metagenomics, Polyhydroxyalkanoate synthase, Shotgun sequencing

Introduction

The increasing challenges associated with plastic waste management and the depletion of natural resources used in plastic production have motivated researchers to develop alternative and environmentally friendly plastics (Amache et al. 2013). One such alternative is polyhydroxyalkanoates (PHAs), which are biodegradable polyesters produced under unfavorable growth and stress conditions by a variety of microorganisms, such as bacteria and archaea. When these microbes are exposed to too much carbon and other nutrient limitations, they create PHAs as water-insoluble inclusions that serve as carbon and energy storage (Chek et al. 2017). The synthesis of PHAs by microorganisms involves several enzymatic reactions, with the key enzyme being PHA synthase. Therefore, this enzyme plays a crucial role in PHA biosynthesis by determining the types of monomers that are added to the chain of the PHA polymer (Zher Neoh et al. 2022).

Currently, four classes of PHA synthases (class I, II, III, and IV) have been reported. These classes constitute a distinct protein family, each of which has a particular substrate specificity, domain, and subunit makeup. Class I PHA synthase is encoded by a single PhaC gene, whereas Class II PHA synthase is encoded by two distinct PhaC genes (PhaC1 and PhaC2). PhaC genes of classes II and I are homodimers (single-unit enzymes), while those of classes III, and IV are heterodimers (consisting of two subunits) that include PhaC-PhaE and PhaC-PhaR, respectively, and both are needed to establish functionality (McCool and Cannon 2001). Class I, III, and IV synthases have been shown to prefer short-chain length monomers (SCLs) with carbon chain lengths between C3 and C5 (Mezzolla et al. 2018). However, class II synthases prefer medium-chain-length (MCL) monomers, including the C6 monomer 3-hydroxyhexanoate (3HHx), which has carbon chain lengths between C6 and C14. Additionally, some PhaCs have been found to be able to add MCL-PHA and SCL-PHA monomers to the PHA polymer chain (Chek et al. 2017). PHA synthase enzymes have remarkable substrate specificity, allowing them to polymerize more than 150 different monomers into various forms of PHA. This broad substrate specificity of PHA synthases is probably due to their low protein sequence similarity. Consequently, the amino acid sequences of PHA synthases can vary significantly, ranging from 8 to 96% similarity (Rehm, 2003). This diversity in sequence makes it challenging to design a single universal primer

set to detect all four classes of PhaC genes. PHA synthases have certain conserved properties at the core structural level. These proteins consist of a catalytic triad, eight highly conserved amino acid residues, and a putative lipase box-like motif in the α/β domain called "G-X-C-X-G" (Zher Neoh et al. 2022). Understanding the different classes of PHA synthases and their substrate specificities is crucial for tailoring PHA production to obtain the desired properties and monomer compositions in biodegradable plastics.

Although PHAs have been known for centuries, the industrial production of these biodegradable plastics remains a challenge due to their high cost (Chen and Jiang 2017). To overcome this hurdle and enable commercialization, researchers have focused on using low-cost renewable feedstock and waste streams to reduce production expenses (Bhuwal et al. 2013). However, the lack of efficient microorganisms capable of converting waste into PHAs has hindered their commercialization, and the cost of PHA production is still 5–10 times greater than that of conventional petrochemical-based plastics (Raza et al. 2018). To address this limitation, researchers have been searching for novel bacterial strains with a high ability to convert low-cost or even cost-free substrates into PHA, while also optimizing the bioprocesses involved in their production. To facilitate the broad use of PHA as a sustainable substitute for conventional plastics, these initiatives aim to increase the productivity and economic viability of PHA production.

To date, most research on the diversity of the PhaC gene and PHA producers has been conducted on pure isolates using culture-dependent methods that compromise the validity of novel uncultured microbial communities. This approach has created a significant knowledge gap about the diversity of PhaC genes and PHA producers in the largely unexplored microbial world. To bridge this gap, metagenomic approaches are essential, as they provide direct access to untapped microbial genomic information. In this context, the Addis Ababa municipal solid waste landfill is a previously unexplored area of research. The study conducted here represents the first attempt to investigate the soil microbial community in this area, and this exploration offered a significant opportunity to discover new PhaC genes from previously unknown microbial genera within the metagenome. Therefore, this study sheds light on the large and undiscovered microbial world in the Addis Ababa municipal solid waste using a shotgun metagenomics sequencing approach to reveal

new possibilities for understanding the diversity of PhaC genes and PHA producers.

Materials and methods

Study area

The research study area was the Addis Ababa municipal solid waste disposal site (Qoshe), which is located in the southwest of the city within the borders of Kolfe Keranio and Nifas Silk-Lafto. The site covers a surface area of 25 hectares and has served as the primary disposal site for solid waste from Addis Ababa city since 1968 (Nademo et al. 2023). From April 20 to April 22, 2023, soil samples from the study area were randomly collected from five distinct sites at depths of 15 to 95 cm at 10 cm intervals at each location in three rounds using aseptic techniques to avoid contamination. The soil samples were collected using sterile polyethylene bags. To obtain representative samples, soil samples from each location and depth were pooled together. The pooled samples were then thoroughly mixed and homogenized in sterilized containers. Finally, the homogenized soil samples were stored at 4 °C until further analysis.

Extraction and purification of total community DNA

Metagenomic DNA was extracted manually according to Zhou et al. (1996) with slight modifications. For manual extraction of metagenomic DNA, soil samples weighing 3 g were mixed with 15 ml of DNA extraction buffer and 100 µl of proteinase K (10 mg/ml) in Falcon tubes. The mixing was carried out by horizontal shaking at 225 rpm for 30 min at 37 °C. After the mixing step, 1.5 ml of 20% SDS was added to the samples, and the samples were incubated in a 65 °C water bath for 2 h with gentle end-to-end inversion every 15 to 20 min. After incubation, the samples were centrifuged at 10,000 rpm for 5 min at room temperature. The supernatants were collected and transferred to 50-ml Falcon tubes. The supernatants were mixed with an equal volume of chloroform-isoamyl alcohol (24:1, vol/vol). The mixtures were centrifuged, and the aqueous phase was recovered. The recovered aqueous phase was precipitated by adding 0.6 volume of isopropanol at room temperature for 1 h. The nucleic acid pellet was obtained by centrifugation at 14,000 rpm for 10 min at room temperature. The pellet was washed with 70% cold ethanol and then resuspended in nucleus-free water to obtain a final volume of 500 µl of purified DNA. Quality control (QC) was performed to guarantee the reliability of the data. Before library preparation, sample quality control was performed using a Thermo Scientific NanoDrop 3300 fluorospectrometer (Thermo Fisher Scientific, Wilmington, DE, USA) and gel electrophoresis. Subsequently, the genomic DNA was randomly sheared into short fragments, end repaired, A-tailed, and ligated with Illumina adapters. The fragments were size selected,

PCR amplified, and purified. The library was quantified with a Qubit fluorometer and real-time PCR, and a bioanalyzer was used for size distribution detection. The quantified libraries were pooled, and shotgun metagenomic sequencing was performed on one lane of a flow cell using a 150 bp paired-end run on a NovaSeq PE150 instrument (Illumina, Tsim Sha Tsui, Hong Kong). The sequenced reads (raw reads) were filtered using Fastp (<https://github.com/OpenGene/fastp>) to remove reads containing adapter contamination, more than 10% uncertain nucleotides, and more than 50% low-quality nucleotides (those with a base quality less than 5).

Bioinformatics and statistical analysis of metagenome data

The soil metagenome sequence was analyzed using the SqueezeMeta pipeline (version 1.5.1) (Tamames and Puente-Sánchez 2019), and both the taxonomic and functional profiles of the metagenomic datasets were inferred. Briefly, in the SqueezeMeta pipeline, the coassembly of the raw reads was performed using Megahit software v1.2.9; short contigs (<150 bps) were removed, and the contig statistics were determined using Prinseq (Schmieder and Edwards 2011); read mapping against contigs was performed using Bowtie2 v2.5.3 (Langmead et al. 2019), and Prodigal v2.6.3 (Hyatt et al. 2010) was used for ORF prediction. In addition to prodigal prediction, a diamond BlastX search (Buchfink et al. 2014) was performed on parts of the contigs where no ORFs were predicted or where the predicted ORFs did not match anything in the taxonomic and functional databases to provide more sensitive ORF detection. Barrnap v0.8 was used for 16 S rRNA gene sequence retrieval (Seemann 2014), which was taxonomically classified using the RDP classifier (Wang et al. 2007). Diamond software v4.6.7 (Buchfink et al. 2014) was used for taxonomic classification of the ORFs against the GenBank nr database (Clark et al. 2016) and functional annotation of ORFs with COG (Huerta-Cepas et al. 2016) and KEGG databases (Kanehisa and Subramaniam 2002). HMMER3 software v 3.3.2 (Eddy 2009) was used for HMM homology searches against the Pfam database (Finn et al. 2016). Pathway prediction for the KEGG (Kanehisa and Subramaniam 2002) and MetaCyc (Caspi et al. 2018) databases was performed using MinPath v1.6 (Ye and Doak 2009). The estimation of coverage and abundance for each gene was carried out by aligning sequence reads to the gene using Bowtie2 v2.5.3 (Langmead et al. 2019). The entire SqueezeMeta project was uploaded to RStudio running the R version 4.3.1 environment using the SQMtools R package functions for further statistical analysis and data visualization.

Bioinformatic prediction of PHA synthase classes

The functional annotation of the metagenomics dataset was carried out using a diamond search against the

KEGG and COG databases and a hmm search against the Pfam database. However, Pfam, KEGG, and COG do not differentiate among PHAC gene classes, as they use single domain models for similarity searches. All ORFs annotated as PhaC genes using the SqueezeMeta pipeline were retrieved from the SQM object via the SQM function after loading the SqueezeMeta project into the R environment (R version 4.3.1). A CD search (Lu et al. 2020) against the TIGRFAMs database of protein families (Haft et al. 2001) was performed to identify the full-length multidomain sequences of PhaC genes and classify them into their respective classes (classes I-III). Clustering of PhaC open reading frames into representative sequences was performed using the CD-HIT protein V4.6 (Li and Godzik 2006). Multiple sequence alignment was performed using Unipro UGENE v1.9 (Okonechnikov et al. 2012) for amino acid sequences of representative phaC genes recovered from the soil metagenome and selected reference PhaC genes from the NCBI database of nonredundant protein sequences (nr). All putative PhaC genes obtained in this study were analyzed for similarity against the NCBI nonredundant protein sequence (nr) database to determine their sequence novelty. The sequences of putative PhaC genes were aligned using the ClustalW program, and their genetic relationships were determined by the maximum likelihood method with the MEGA11 program (Tamura et al. 2021). Bootstrap replicates of data were analyzed to assess the level of statistical support for branches in the phylogeny.

Results

Overview of the metagenomic data

A combined sample from different sites and depths was used for DNA extraction and sequencing. In general, Illumina sequencing (Hong Kong, China) using Nova-SeqPE150 (Novogene, Hong Kong, China) generated an average of 16,860 Mbp raw bases with GC content

determined to be 65.40%. After quality control filtering, 16,810 Mbp clean bases were retained for further analysis. The assembly resulted in a total of 1,009,810 contigs. The longest contig was 463,223 bases in length, and the N50 parameter was 1,527 bases. A total of 1,626,507 ORFs were predicted for functional annotation, of which approximately 54% were annotated against the KEGG database, 74% against the COG database, and 42% against the Pfam database. The data were evaluated for the distribution of base call quality, and the percentages of >Q20 and >Q30 bases were 97.47% and 93.47%, respectively. Sequencing statistics are shown in Table 1.

Bacteria and PHA producer community composition

The taxonomic composition analysis of the Qoshe landfill revealed that the domain Bacteria dominated the microbial community, accounting for 92% of the total microbiome. The remaining community members consisted of Archaea, Eukaryota, and viruses, which were considered insignificant. Interestingly, over 5% of the genes in the municipal solid waste sample remained unclassified. At the phylum level, the most abundant bacterial groups were Actinomycetota (56% of total community), Proteobacteria (23% of total community), unclassified bacteria (2.9% of total community), Chloroflexota (2.6% of total community), and Bacillota (2.1% of total community) (Fig. 1a). Further analysis at the genus level showed that the metagenome was clearly dominated by *Mycobacterium* (27% of total community), *Methylocaldum* (5% of total community), *Saccharomonospora* (4% of total community), and unclassified Actinomycetia (1% of total community) (Fig. 1b).

The metagenomic analysis of the Qoshe landfill ecosystem revealed a remarkable diversity of PHA-producing microorganisms. Within the bacterial community, 235 ORFs were found to be related to the PhaC enzyme, accounting for approximately 0.016% of the total bacterial ORFs. The most abundant PhaC-producing bacteria belonged to the phyla Proteobacteria (0.0097% of total bacteria), Actinobacteria (0.00192% of total bacteria), Chloroflexi (0.00034% of total bacteria), and unclassified bacteria (0.00014% of total bacteria) (Fig. 1c). Notably, the majority of PhaC ORFs were attributed to genera such as *Thiomonas* (0.0021% of total bacteria), *Methylocaldum* (0.0017% of total bacteria), unclassified Proteobacteria (0.0015% of total bacteria), and unclassified Actinomycetia (0.0013% of total bacteria) (Fig. 1d), highlighting the vast untapped potential for discovering novel PHA-producing microorganisms within the Qoshe landfill ecosystem.

Table 1 Shotgun metagenomics sequencing statistics of municipal solid waste soil sample

Parameters	Qoshe
Number of Raw Bases (Mbp)	16,860
Clean Base (Mbp)	16,810
Library insert size (bp)	350
Length of single read (bp)	150
GC content (%)	65.4
Number of contigs	1,009,810
Length of longest contig (bp)	463,223
Length of shortest contig (bp)	200
N50 (bp)	1,527
N90 (bp)	577
Mapping (%)	82.6
Number of ORFs	1,626,507
Number of bacterial ORFs	1,493,877

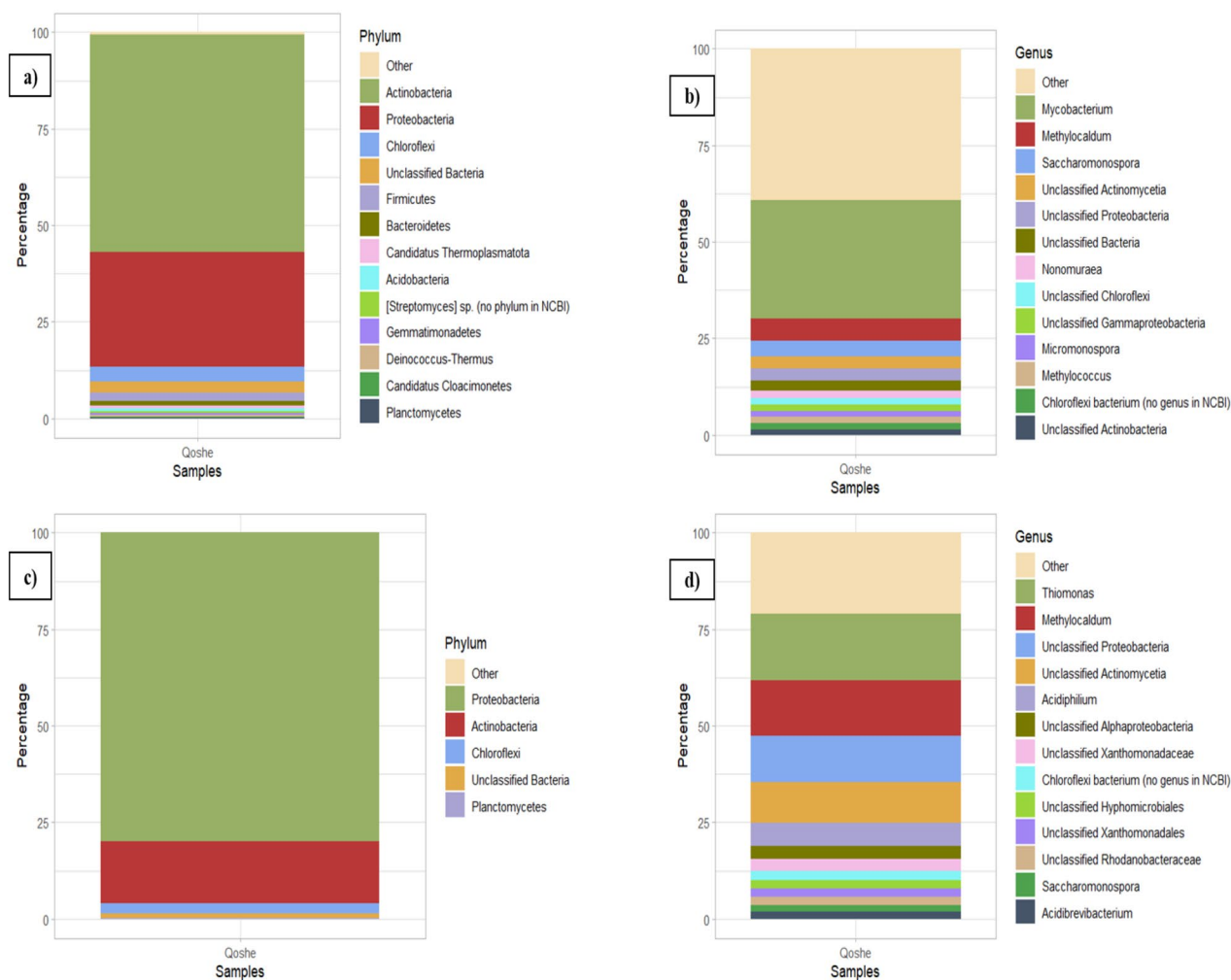


Fig. 1 Taxonomic composition and percent relative abundance of bacterial community and PHA producing taxa detected in the municipal solid waste soil metagenome. Bacterial community at phylum (a) and genus (b) levels, as well as PHA producing taxa at phylum (c) and genus (d) levels

Functional and taxonomic profiling of PhaC genes in municipal solid waste soil

Metagenomic functional annotation revealed 235 putative PhaC open reading frames (ORFs) in the Addis Ababa Municipal Solid Waste Soil Metagenome. To classify the putative PHA synthase genes into their respective classes, a CD search was performed against the TIGRFAMs database of protein families. The TIGRFAM models successfully differentiated the sequences of the three classes of PhaC genes (Additional file 3: Table S2). Of the initial 235 PhaC ORFs, 21 were excluded from subsequent analysis because they did not show similarity to PhaC genes in the TIGRFAM database. A CD search against the TIGRFAM database revealed that, of the remaining 214 PhaC ORFs, 122 were associated with class I PhaC genes, 9 with class II PhaC genes, and 83 with class III PhaC genes. The class I PhaC genes were found to be the most abundant, followed by the class III PhaC genes in the overall dataset.

A comparison of the amino acid sequences of all 214 PhaC ORFs (class I-III PhaC genes) revealed that they could be classified into 14 genetic groups based on amino acid sequence similarity (Additional file 2). This grouping was correlated with the maximum likelihood phylogenetic tree (Additional file 1), where genes designated the same GG clustered closely together. The protein sequence identities of the 214 PhaC ORFs against the NCBI database of nonredundant protein sequences (nr) ranged from 48 to 100% (Additional file 3 table S1), with 46.5% of PhaC ORFs showing >90% identity to the protein database (Fig. 2a). Although many putative PhaC ORFs showed high identity to reference proteins, the amino acid sequences of most PhaC ORFs exhibited significant variation from each other, with 45.32% of the ORFs having less than 50% similarity (Fig. 2b). Furthermore, the amino acid sequences of the representative PhaC genes were aligned with other commonly known PhaC amino acid sequences in the database. The results

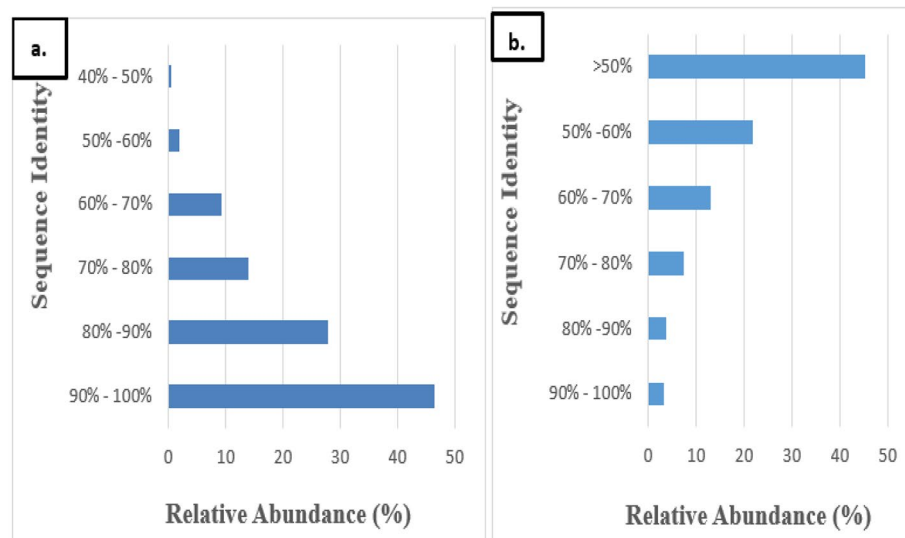


Fig. 2 (a) Putative PhaC gene amino acid sequence identity with the NCBI nonredundant protein sequence (nr) database; and (b) amino acid sequence similarity among putative PhaC genes identified in this study

showed that all representative PhaC genes contained a lipase-box-like motif “G-X-C-X-G” except for GG13 and GG14. Conserved residues were also found among all the sequences in the shaded regions (Fig. 3).

Class I PHA synthase genes

In the soil metagenomes, unclassified Burkholderiales (32.24% of the PhaC genes), unclassified Alphaproteobacteria (22.67%), Mycobacteriaceae (17.02%), Acetobacteraceae (10.41%), and Hyphomicrobiaceae (6.5%) were the most dominant families contributing to class I PhaC genes (Fig. 4a).

The protein sequence identities of class I PhaC genes ranged from 64.13 to 100% (Additional file 3 Table S1). Specifically, when examining the top hits, the percentage identities to unclassified Burkholderiales, Acetobacteraceae, and unclassified Alphaproteobacteria ranged from 65.95 to 98.92%, 99.67–100%, and 69.33–98.27%, respectively. The class I PhaC genes also exhibited similarities of 73.21–100% with genes found in Hyphomicrobiaceae. Interestingly, despite the well-known reputation of members of the Rhodocyclales as short-chain-length PHA producers, a BLAST search conducted on 7 March 2024 revealed that most of the detected class I PhaC genes affiliated with these taxa exhibited novel sequences. Their identities in the database ranged from 60.7 to 93.07%. On the other hand, genes detected from taxa that are less commonly associated with PHA production, such as Stappiaceae and Mycobacteriaceae, showed greater similarity to database sequences, with identity matches of 84% and 79.41%, respectively.

Class II PHA synthase genes

The functional annotation revealed that the class II PhaC genes were the least abundant among the identified PhaC genes in the overall dataset (Additional file 3 Table S2). According to the familywise distribution of the metagenomics data, Pseudomonadaceae, unclassified Hyphomicrobiales, unclassified Acidimicrobiales, unclassified Alphaproteobacteria, and Methylocystaceae were the dominant contributors to class II PhaC gene metagenome hits (Fig. 4b). In terms of abundance, Pseudomonadaceae represented the majority of genes in the soil metagenome (66.93% of the PhaC genes), followed by unclassified Hyphomicrobiales (8.95%), Acidimicrobiales (7.78%), unclassified Alphaproteobacteria (2.72%), and Methylocystaceae (2.33%). The detected class II PhaC genes exhibited the highest identity matches (97.52–99.82%) with *Pseudomonas*, which are well-known medium-chain length PHA producers. However, relatively lower percentage identities were observed for PhaC genes related to unclassified Acidimicrobiales and unclassified Hyphomicrobiales, with values of 69.08% and 55.77%, respectively (Additional file 3 Table S1). The class II PhaC genes showed the closest identity (81.48%) to the *Methylocystis* gene.

Class III PHA synthase genes

Methylococcaceae dominated the contribution of class III PhaC genes in the soil metagenome, making up 49.65% of the PhaC genes. Other bacterial families, including Xanthomonadaceae, *Chloroflexi* bacterium, Acetobacteraceae, and Gammaproteobacteria, also made moderate contributions to class III PhaC genes in the soil metagenome (15.32%, 8.67%, 4.91%, and 4.33%, respectively)

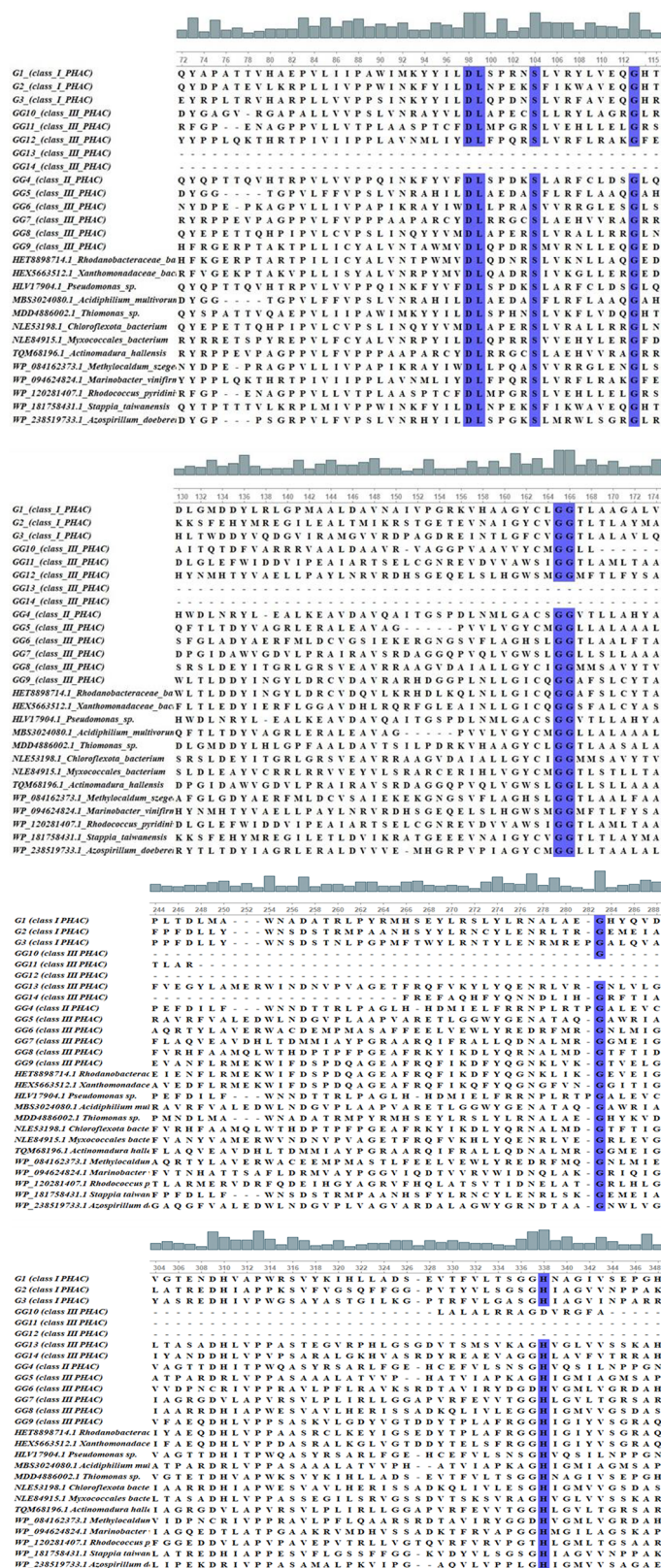


Fig. 3 Multiple alignment of amino acid sequences from GG1 to GG14. MSA was performed with Class I-III PhaC genes from known PHA producers and PhaC obtained from this study. Amino acid residues highlighted in blue were found to be conserved among all PhaC genes

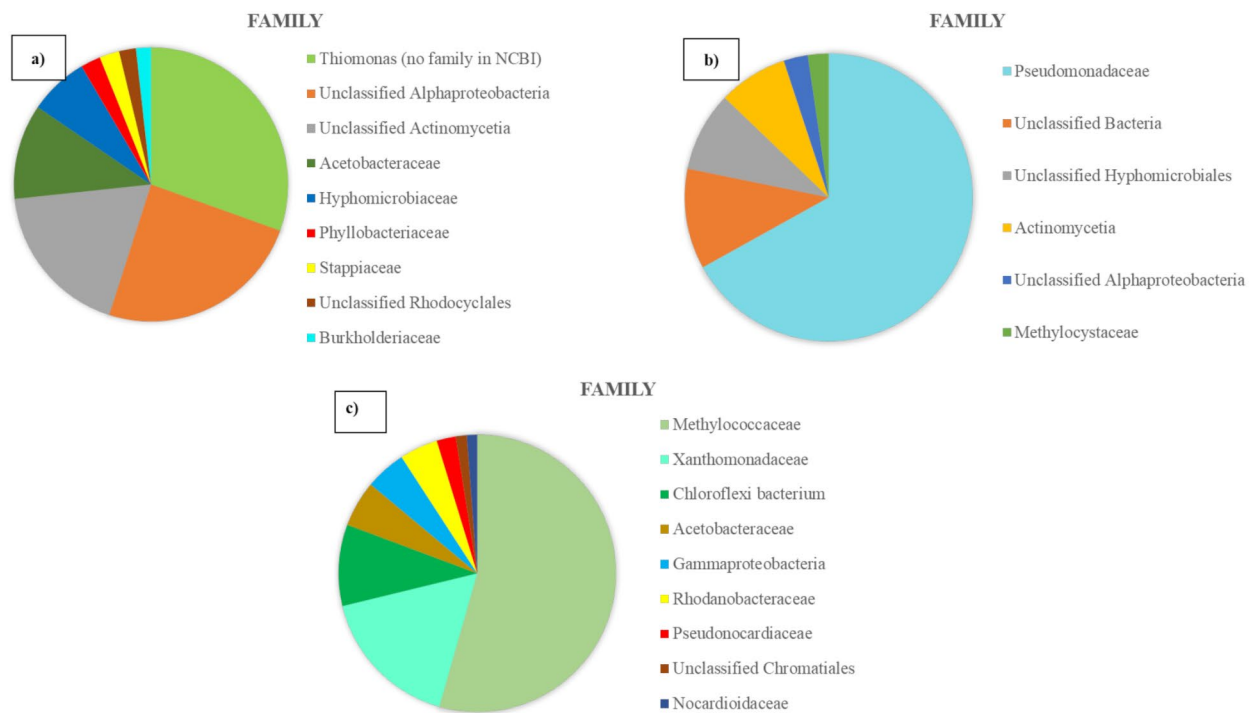


Fig. 4 Taxonomic distribution of class I (a), II (b), and III (c) PhaC genes detected in municipal solid waste soil metagenomes

(Fig. 4c). The protein sequence identities of the class III PhaC genes varied between 48% and 100%. The class III PhaC genes exhibited a percentage identity of 85.07–100% with *Methylococcaceae*, a well-known family that produces PHA. Similarly, the identity range for *Xanthomonadaceae*, another recognized PHA producer, was 76.07–100%. The identity range with *Chloroflexota* bacterium was 61.24–100%, while with *Gammaproteobacteria*, it was 70.62–99.44%. In the soil metagenome, the average percentage identities of class III PhaC genes to *Acetobacteraceae* and *Rhodanobacteraceae* ranged from 61.24 to 100% and 78.59–90.65%, respectively.

Discussion

Landfill areas are recognized as significant microbial reservoirs with potential for bioprospecting (Song et al. 2015). Recent advances in genomics provide an opportunity to extensively study these valuable microbial resources within landfill areas for a variety of applications. The metagenomic analysis of the microbial community in the Qoshe landfill provides novel insights, as this environment has not been extensively studied before in terms of its microbial composition, particularly regarding the diversity of PHA producers. This landfill served as a repository for a diverse array of waste streams from all areas of Addis Ababa City for over 50 years. Due to its chemical composition, this area can be considered a promising target area for exploring novel PHA

producers. This is because, as reported before, environments that are naturally or accidentally exposed to high organic carbon or growth-limiting conditions can serve as potential sources of PHA-producing microbes (Singh Saharan et al. 2014). Desta (2022) reported that this open landfill contains approximately 60% organic waste, 15% recyclable waste, and 25% other waste such as paper, rubber/plastics, wood, bone, textiles, metals, glass, and non-combustible stone. The Qoshe landfill is also known to produce and emit large quantities of methane gas to the atmosphere (Jigar et al. 2014). This high methane concentration, along with the elevated concentration of organic carbon within the landfill, can enrich microbial communities capable of producing PHAs. Generally, exploration of PHA producer microbes from landfill areas is of significant importance, especially when considering the potential to couple waste management with PHA production for the development of a circular economy. Herrera et al. (2023) examined plastic-polluted sites, landfills, and oil-polluted sites as sources of microorganisms with the potential ability to convert plastic substrates to PHA. However, collecting samples from landfill environments is challenging due to the high physical and chemical heterogeneity of the waste, as well as the safety concerns associated with accessing and sampling these complex waste management systems. To overcome these challenges, the present study collected a large number of samples over three consecutive days to ensure a thorough

representation of the landfill's microbial diversity for the metagenomics study. So far, most studies on the diversity of PHA producers and PhaC genes rely on PCR amplification of PhaC genes from environmental samples (Foong et al. 2014; Yang et al. 2013; Cheema et al. 2012). This approach has limited the discovery of novel PHA-producing taxa due to the constraints of existing primers, which do not adequately capture the sequence diversity of environmental homologs (Thi and Vahlis 2017). Overall, this study was designed to investigate the diversity of the PHA-producing bacterial community and the PhaC genes in municipal solid waste samples using the shotgun Illumina high-throughput sequencing approach.

The metagenomic analysis revealed the microbial community's taxonomic diversity, with key bacterial taxa such as Actinomycetota, Proteobacteria, and Chloroflexota dominating. This result was consistent with the findings by Selvarajan et al. (2022), who reported Actinomycetota as the dominant taxa in informal dump sites. Notably, the discovery of a substantial proportion (over 5%) of unclassified genes indicates the existence of undiscovered novel microbial diversity, highlighting the unexplored biotechnological and ecological potential of the Qoshe landfill. Additionally, detailed metagenomic analysis revealed a notable diversity of PHA-producing microorganisms, with the relative abundance of PhaC open reading frames (ORFs) accounting for 0.016% of the total bacterial ORFs. Yasuda et al. (2021) found more PhaC genes (0.171% of all bacterial genes) in the activated sludge of a leachate treatment plant, which was relatively higher than our result. Our study may have underestimated the real abundance of PHA producers due to the physical heterogeneity of the solid waste compared to leachate. The study by Foong et al. (2018) also listed a total of 8 microbial PHA-producing phyla. Our study, on the other hand, found 11 PHA-producing phyla, demonstrating the diversity of PHA producers in Qoshe landfills. According to Ospina-Betancourth et al. (2022), treating sludge and wastewater from the yeast production industry resulted in an increase in the relative abundance of PHA producer bacteria from 37% to 78.52%. However, the present study subjected a raw waste soil sample to DNA extraction and sequencing without any pretreatment, potentially contributing to the low abundance of PHA producers. PhaC-producing bacteria primarily belong to the phyla Proteobacteria, Actinobacteria, Chloroflexi, and unclassified bacteria, with the majority of PhaC ORFs attributed to the genera *Thiomonas*, *Methylocaldum*, unclassified Proteobacteria, and unclassified Actinomycetia (Fig. 1a). These results are largely consistent with the literature on known PHA producers. Foong et al. (2018) analyzed the mangrove soil microbial community and revealed that approximately 86% of the genera producing PHA were derived from the phylum Proteobacteria. The

accumulation of PHA as an energy storage polymer is well established in members of the phylum Actinomycetota and Proteobacteria (Foong et al. 2014; Ciesielski et al. 2010). However, the association of the Chloroflexota phylum with PHA production has been studied less extensively than that of Pseudomonadota and Actinobacteria. Interestingly, the finding that a substantial number of these PhaC ORFs exhibit low sequence similarity to known PhaC enzymes in databases may suggest the presence of unique microbial metabolic pathways, warranting further exploration and potential biotechnological applications.

Among the PhaC genes obtained in this study, class I PhaC was the most dominant class, followed by class III PhaC genes. Most of the class I PhaC genes were affiliated with members of the Burkholderiales order (Fig. 4a), suggesting their relevance among PHA producers in landfill areas. The most representative genus of the Burkholderiales order was *Thiomonas*. The discovery of *Thiomonas* as the most dominant genus related to PhaC genes is intriguing, particularly considering the limited literature on PHA production in this genus. Although members of *Thiomonas* are less commonly associated with PHA production, metagenomics hits for these taxa had identities to reference sequences ranging from 78.32 to 98.92%, which indicates less novelty of the sequence. Previous studies have highlighted *Thiomonas* as a promising organism for various bioremediation applications, such as biofiltration of industrial air streams contaminated with CS₂ and H₂S (Pol et al. 2007) and treatment of arsenic-contaminated environments, such as mine tailings and drainage water (Teoh et al. 2017). *Thiomonas* species can exhibit chemolithoautotrophic growth using sulfur ores, arsenopyrite, or H₂S as energy sources. Additionally, they have demonstrated the ability to mobilize copper, zinc, and uranium from ore mixtures and exhibit high levels of resistance to cobalt, nickel, and zinc (Kelly et al. 2007). These characteristics of *Thiomonas*, including its ability to remove environmental pollutants, its ability to grow at low pH, and its resistance to various metals and contaminants, make it a valuable candidate for coupling waste remediation with PHA production. In the present study, unclassified *Mycobacterium* were also identified as one of the main contributors to class I PhaC genes. *Mycobacterium* species are known for their ability to degrade complex organic compounds and are widely known to cause diseases in warm-blooded animals (Gunasingam 2022). However, they are relatively understudied in the context of PHA production. A study by Huang et al. (2012); Purohit et al. (2008) showed that members of the *Mycobacterium* genus are capable of accumulating scl PHA.

Furthermore, a metagenomic investigation of Addis Ababa municipal solid waste revealed the presence of

class II PhaC genes. These genes were predominantly related to the Pseudomonadaceae family, as shown in Fig. 4. Within the Pseudomonadaceae family, the *Pseudomonas* genus was identified as a prominent contributor to class II PhaC genes. *Pseudomonas* species are well known for their ability to produce medium chain length (mcl) PHA, as supported by previous studies by Adebayo Oyewole et al. (2024); Kanavaki et al. (2021). The unclassified Hyphomicrobiales and Acidimicrobiales orders were found to contribute significant proportions of class II PhaC genes. However, knowledge regarding their ability to accumulate mcl PHAs is currently limited. The sequence identity of the class II PhaC genes with unclassified Hyphomicrobiales was 55.77%, while that with Acidimicrobiales ranged from 66.99 to 69.08%. This result shows the high novelty of the sequence and suggests that the genes most likely belong to a new microbial genus within these two orders. Members of the Methylocystaceae family also made a moderate contribution to class II PHA genes. The genus *Methylocystis*, which is a representative genus of Methylocystaceae in this study, has been described in various studies on PHA production (Cheng 2020; Fergala et al. 2018; Levett et al. 2016).

Metagenomic analysis revealed a high taxonomic abundance of class III PhaC genes from members of the Methylococcaceae family. Further analysis at the genus level revealed that most of the class III PhaC genes were affiliated with the genus *Methylocaldum*. It is well established that members of the genus *Methylocaldum* can accumulate PHA as an energy storage polymer, as supported by previous studies (Chavan et al. 2021; Bhatia et al. 2021). *Methylocaldum* bacteria can utilize methane as a carbon source and can thrive in high-temperature environments (Luangthongkam et al. 2019; Levett et al. 2016). This finding highlights the possibility of using methane-rich environments, such as municipal solid waste, as a cheap carbon source for low-cost thermophilic PHA production by *Methylocaldum* species. Metagenomic analysis also revealed the significant presence of class III phaC genes contributed by the Xanthomonadaceae family and the *Chloroflexi* bacterium, as shown in Fig. 4d. Notably, these two taxa have received relatively less attention in the context of PHA production, and their involvement in this process remains understudied. The analysis of sequence novelty for class III phaC genes affiliated with the *Chloroflexi* bacterium and the Xanthomonadaceae family revealed a significant degree of variation. Specifically, the sequence identities of the PhaC genes ranged from 61.24 to 100% for the *Chloroflexi* bacterium and from 76.07 to 100% for the Xanthomonadaceae family, indicating substantial variation in the novelty of the sequence within these two taxa.

In the present study, the amino acid sequences of putative PhaC genes (class I-III PhaC genes) were compared

with each other and with reference proteins in the database. The analysis revealed that 46.5% of the putative PhaC genes shared more than 90% identity with the reference proteins. Interestingly, the sequence identity among the putative PhaC genes was found to be very low. Approximately 45.33% of the putative PhaC genes exhibited less than 50% identity. Analysis of amino acid sequences of the PhaC genes revealed an unexpected level of sequence variation among the genes within the same PhaC gene classes, which contrasts with the expected high similarity of genes within the same class. Especially, for GG1, GG7, and GG10 different PhaC gene classes were classified into the same genetic groups (Additional file 2 Table S1), suggesting an exceptional level of diversity within the established PhaC class definitions. This exceptional variation observed within the same PhaC classes could potentially indicate the presence of previously unidentified or atypical PhaC gene variants. Alternatively, it may suggest that the current PhaC class definitions need to be refined to better account for the observed sequence diversity and more accurately capture the true extent of variations in these genes. Overall, the high variation between or within PhaC gene classes indicates the presence of extreme diversity in PhaC genes within the metagenome dataset.

All representative putative PhaC genes obtained in this study were found to have conserved amino acid residues of a lipase box (G-X-C/S-X-G), except for GG13 and GG14, which are essential active sites in PHA synthase (Fig. 3). In the cases of GG10 and GG11, the first glycine of the lipase box was substituted with alanine and valine, respectively. Similarly, Foong et al. (2014) analyzed the amino acid sequences of 55 PhaC genes and noted that the first glycine of the lipase box was replaced by serine (S) and alanine (A) in two PhaC GGs. This substitution alters the amino acid composition of the lipase box motif and may have implications for modifying the function or activity of the PHA synthase enzyme. Furthermore, metagenomic analysis of the Addis Ababa Municipal Solid Waste Soil also revealed the presence of a diverse community of potential PHA producers. These findings not only expand our knowledge of PHA-producing taxa but also provide opportunities for further exploration and utilization of these organisms for sustainable PHA production from waste resources.

Conclusion

Overall, by expanding the scope beyond culturable bacteria and employing a metagenomic approach, this study offers a more comprehensive understanding of the diversity of PHA producers and PhaC genes in the Addis Ababa municipal solid waste disposal site. This knowledge can contribute to future research and applications related to sustainable PHA production, waste

management strategies, and the utilization of microbial resources. These findings contribute to our understanding of the role of the microbial community in PHA production and highlight potential targets for further investigation and exploitation of these organisms for sustainable PHA production from waste resources. Further studies, including the isolation and characterization of specific strains, are necessary to confirm their PHA production capabilities and explore their biotechnological potential.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13213-024-01778-3>.

Additional file 1: Figure S1: Protein maximum likelihood phylogenetic tree of PHA synthase (PhaC) genes. All putative PhaCs obtained from metagenomic studies were subjected to phylogenetic classification

Additional file 2: PhaC gene clusters created using CD-HIT protein software

Additional file 3: Table S1: Genetic groups and closest organism matches of phaC genes; Table S2: Results of the CD search against the TIGRFAM database

Acknowledgements

We thank Addis Ababa Science and Technology University for providing funding to carry out this research. The authors also thank the Department of Biotechnology for providing the necessary infrastructure to conduct this research.

Author contributions

Conceptualization, Zuriash Mamo and Mesfin Tafesse; methodology, Zuriash Mamo and Mesfin Tafesse; formal analysis, Zuriash Mamo, Sewunet Abera and Mesfin Tafesse; investigation, data curation, writing manuscript and editing, Zuriash Mamo and Mesfin Tafesse; supervision, Mesfin Tafesse; project administration, Mesfin Tafesse. All authors have read and agreed to the published version of the manuscript.

Funding

Addis Ababa Science and Technology University financially supported this research, which was used for sample collection, procurement of laboratory chemicals and reagents, and payment for some laboratory works.

Data availability

The metagenomic raw reads obtained in this study have been deposited in the Sequence Read Archive (SRA) at NCBI. The corresponding accession number for the metagenomic reads is PRJNA1103368. The sample name associated with these reads was "Qoshe soil sample."

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that there are no conflicts of interest with respect to the publication of this article.

Author details

¹Department of Biotechnology, College of Natural and Applied Sciences, Addis Ababa Science and Technology University, Addis Ababa.P.O. Box 16417, Ethiopia

²Biotechnology and Bioprocess Center of Excellence, Addis Ababa Science and Technology University, Addis Ababa.P.O. Box 16417, Ethiopia

³Ethiopian Institute of Agricultural Research (EIAR), Addis Ababa.P.O. Box 2003, Ethiopia

⁴Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), P.O. Box 50, Wageningen 6700 AB, The Netherlands

⁵Institute of Biology, Leiden University, P.O. Box 9500, Leiden 2300 RA, The Netherlands

Received: 8 May 2024 / Accepted: 10 September 2024

Published online: 01 October 2024

References

- Adebayo Oyewole O, Usman Abdulmalik S, Onozasi Abubakar A, Ishaku Chimbejuwo K, Dorcas Obafemi Y, Oyegbile B, Abioye P, Adeniyi OD, O, Egwin C (2024) E. Production of polyhydroxyalkanoate (pha) by pseudomonas aeruginosa (ol405443) using agrowastes as carbon source. *Cleaner Materials*, 11(October 2023), 100217. <https://doi.org/10.1016/j.clema.2024.100217>
- Amache R, Sukan A, Safari M, Roy I, Keshavarz T (2013) Advances in PHAs production. *Chem Eng Trans* 32:931–936. <https://doi.org/10.3303/CET1332156>
- Bhatia SK, Otari SV, Jeon JM, Gurav R, Choi YK, Bhatia RK, Pugazhendhi A, Kumar V, Rajesh Banu J, Yoon JJ, Choi KY, Yang YH (2021) Biowaste-to-bioplastic (polyhydroxyalkanoates): Conversion technologies, strategies, challenges, and perspective. *Bioresour Technol* 326(November 2020):124733. <https://doi.org/10.1016/j.biortech.2021.124733>
- Bhuwal AK, Singh G, Aggarwal NK, Goyal V, Yadav A (2013) Isolation and screening of polyhydroxyalkanoates producing bacteria from pulp, paper, and cardboard industry wastes. *Int J Biomaterials* 2013. <https://doi.org/10.1155/2013/752821>
- Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59–60. <https://doi.org/10.1038/nmeth.3176>
- Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD (2018) The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* 46(D1):D633–D639. <https://doi.org/10.1093/nar/gkx935>
- Chavan S, Yadav B, Tyagi RD, Drogui P (2021) A review on production of polyhydroxyalkanoate (PHA) biopolyesters by thermophilic microbes using waste feedstocks. *Bioresour Technol* 341(September):125900. <https://doi.org/10.1016/j.biortech.2021.125900>
- Cheema S, Bassas-Galia M, Sarma PM, Lal B, Arias S (2012) Exploiting metagenomic diversity for novel polyhydroxyalkanoate synthases: production of a terpolymer poly(3-hydroxybutyrate-co-3-hydroxyhexanoate-co-3-hydroxyoctanoate) with a recombinant *Pseudomonas putida* strain. *Bioresour Technol* 103(1):322–328. <https://doi.org/10.1016/j.biortech.2011.09.098>
- Chek MF, Kim SY, Mori T, Arsal H, Samian MR, Sudesh K, Hakoshima T (2017) Structure of polyhydroxyalkanoate (PHA) synthase PhaC from *Chromobacterium* sp. USM2, producing biodegradable plastics. *Sci Rep* 7(1):1–15. <https://doi.org/10.1038/s41598-017-05509-4>
- Chen GQ, Jiang XR (2017) Engineering bacteria for enhanced polyhydroxyalkanoates (PHA) biosynthesis. *Synth Syst Biotechnol* 2(3):192–197. <https://doi.org/10.1016/j.synbio.2017.09.001>
- Cheng K (2020) Optimizing the Methanotrophic Production of Polyhydroxyalkanoates Using Mixed Microbial Bacteria Cultures. *Civil and Environmental Engineering Theses and Dissertations*. https://scholar.smu.edu/engineering_civil_etds/11
- Ciesielski S, Pokoj T, Klimiuk E (2010) Cultivation-dependent and -independent characterization of microbial community producing polyhydroxyalkanoates from raw glycerol. *J Microbiol Biotechnol* 20(5):853–861. <https://doi.org/10.4014/jmb.0909.09038>
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. *Nucleic Acids Res* 44(D1):D67–D72. <https://doi.org/10.1093/nar/gkv1276>
- Desta H (2022) An overview of solid waste management systems in the city administration of Addis Ababa: past to present. *LSE Cities* 01(May):1–29. <https://www.lse.ac.uk/Cities/Assets/Documents/RRR/RRR-field-report-01-Addis-Ababa.pdf>
- Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. *Genome Informatics. International Conference on Genome Informatics*, 23(1), 205–211. https://doi.org/10.1142/9781848165632_0019
- Fergala A, Alsayed A, Khattab S, Ramirez M, Eldyasti A (2018) Development of methane-utilizing mixed cultures for the production of

- polyhydroxyalkanoates (PHAs) from anaerobic Digester Sludge. *Environ Sci Technol* 52(21):12376–12387. <https://doi.org/10.1021/acs.est.8b04142>
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Foong CP, Lau N, Deguchi S, Toyofuku T, Taylor TD, Sudesh K (2014) *Whole genome amplification approach reveals novel polyhydroxyalkanoate synthases (PhaCs) from Japan Trench and Nankai Trough seawater*. 1–11. <https://doi.org/10.1186/s12866-014-0318-z>
- Foong CP, Lakshmanan M, Abe H, Taylor TD, Foong SY, Sudesh K (2018) A novel and wide substrate specific polyhydroxyalkanoate (PHA) synthase from unculturable bacteria found in mangrove soil. *J Polym Res* 25(1). <https://doi.org/10.1007/s10965-017-1403-4>
- Gunasingam N (2022) Morphology and pathological characteristics of Mycobacteria pathologic characteristics of the mycobacteria genus. *Mycobact Dis* 12(1000005):1. <https://doi.org/10.35248/2161-1068.22.54.005>
- Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* 29(1):41–43. <https://doi.org/10.1093/nar/29.1.41>
- Herrera DAG, Mojicevic M, Pantelic B, Joshi A, Collins C, Batista M, Torres C, Freitas F, Murray P, Nikodinovic-Runic J, Fournet B, M (2023) Exploring microorganisms from Plastic-Polluted sites: unveiling plastic degradation and PHA production potential. *Microorganisms* 11(12). <https://doi.org/10.3390/microorganisms11122914>
- Huang YT, Chen PL, Semblante GU, You SJ (2012) Detection of polyhydroxyalkanoate-accumulating bacteria from domestic wastewater treatment plant using highly sensitive PCR primers. *J Microbiol Biotechnol* 22(8):1141–1147. <https://doi.org/10.4014/jmb.1111.11040>
- Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, Von Mering C, Bork P (2016) EGGNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44(D1):D286–D293. <https://doi.org/10.1093/nar/gkv1248>
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Integrated Nr Database in protein annotation system and its localization. *Nat Commun* 6(1):1–8. <http://dx.doi.org/10.1016/B978-0-12-407863-5.00023-X%5Cnhttp://www.nature.com/doi/10.1038/ismej.2009.79%5Cnhttp://www.nature.com/doi/10.1038/nature09916%5Cnhttp://dx.doi.org/10.1038/srep25982%5Cnhttp://dx.doi.org/10.1038/ismej.2010.144%5Cnhttp>
- Jigar E, Bairu A, Jigar E, Bairu A, Gesesew A (2014) Application of IPCC model for estimation of methane from municipal solid waste landfill. *J Environ Sci Water Resour* 3(4):52–58
- Kanavaki I, Drakonaki A, Geladas ED, Spyros A, Xie H, Tsiotis G (2021) Polyhydroxyalkanoate (Pha) production in pseudomonas sp. phdV1 strain grown on phenol as carbon sources. *Microorganisms* 9(8). <https://doi.org/10.3390/microorganisms9081636>
- Kanehisa M, Subramaniam (2002) The KEGG database. *Novartis Foundation Symposium*, 247, 91–103. <https://doi.org/10.1002/0470857897.ch8>
- Kelly DP, Uchino Y, Huber H, Amils R, Wood AP (2007) Reassessment of the phylogenetic relationships of *Thiomonas cuprina*. *Int J Syst Evol Microbiol* 57(11):2720–2724. <https://doi.org/10.1099/ijs.0.65537-0>
- Langmead B, Wilks C, Antonescu V, Charles R (2019) Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35(3):421–432. <https://doi.org/10.1093/bioinformatics/bty648>
- Levett I, Birkett G, Davies N, Bell A, Langford A, Laycock B, Lant P, Pratt S (2016) Techno-economic assessment of poly-3-hydroxybutyrate (PHB) production from methane - the case for thermophilic bioprocessing. *J Environ Chem Eng* 4(4):3724–3733. <https://doi.org/10.1016/j.jece.2016.07.033>
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A (2020) CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Res* 48(D1), D265–D268. <https://doi.org/10.1093/nar/gkz991>
- Luangthongkam P, Laycock B, Evans P, Lant P, Pratt S (2019) Thermophilic production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by a mixed methane-utilizing culture. *New Biotechnol* 53(June):49–56. <https://doi.org/10.1016/j.nbt.2019.06.008>
- McCool GJ, Cannon MC (2001) PhaC and PhaR are required for polyhydroxyalkanoic acid synthase activity in *Bacillus megaterium*. *J Bacteriol* 183(14):4235–4243. <https://doi.org/10.1128/JB.183.14.4235-4243.2001>
- Mezzolla V, D'Urso OF, Poltronieri P (2018) Role of PhaC type I and type II enzymes during PHA biosynthesis. *Polymers* 10(8):1–12. <https://doi.org/10.3390/polym10080910>
- Nademo ZM, Shibeshi NT, Gameda MT (2023) Isolation and screening of low-density polyethylene (LDPE) bags degrading bacteria from Addis Ababa municipal solid waste disposal site Koshe. *Ann Microbiol* 73(1). <https://doi.org/10.1186/s13213-023-01711-0>
- Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, Grehov G, Kandrov OG, Rasputin D, Syabro K, M., Tleukenov T (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28(8):1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>
- Ospina-Betancourth C, Echeverri S, Rodriguez-Gonzalez C, Wist J, Combariza MY, Sanabria J (2022) Enhancement of PHA production by a mixed Microbial Culture using VFA obtained from the fermentation of Wastewater from yeast industry. *Fermentation* 8(4). <https://doi.org/10.3390/fermentation8040180>
- Pol A, Van Der Drift C, Op D, Camp HJM (2007) Isolation of a carbon disulfide utilizing *Thiomonas* sp. and its application in a biotrickling filter. *Appl Microbiol Biotechnol* 74(2):439–446. <https://doi.org/10.1007/s00253-006-0663-4>
- Purohit H, Cheema S, Lal S, Raut C, Kalia V (2008) In search of drug targets for *Mycobacterium tuberculosis*. *Infect Disorders - Drug Targets* 7(3):245–250. <https://doi.org/10.2174/187152607782110068>
- Raza ZA, Abid S, Banat IM (2018) Polyhydroxyalkanoates: Characteristics, production, recent developments and applications. *International Biodeterioration and Biodegradation*, 126(September 2017), 45–56. <https://doi.org/10.1016/j.ibiod.2017.10.001>
- Rehm, B. H. (2003). Polyester synthases: natural catalysts for plastics. *Biochemical Journal*, 376(1), 15–33
- Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Selvarajan R, Ogola H, Kalu CM, Sibanda T, Obize C (2022) Bacterial communities in Informal dump sites: a Rich source of Unique Diversity and Functional Potential for Bioremediation Applications. *Appl Sci (Switzerland)* 12(24). <https://doi.org/10.3390/app122412862>
- Singh Saharan B, Grewal A, Kumar P (2014) Biotechnological Production of polyhydroxyalkanoates: a review on trends and latest developments. *Chin J Biology* 2014:1–18. <https://doi.org/10.1155/2014/802984>
- Song L, Wang Y, Tang W, Lei Y (2015) Bacterial community diversity in municipal waste landfill sites. 266. <https://doi.org/10.1007/s00253-015-6633-y>
- Tamames J, Puente-Sánchez F (2019) SqueezeMeta: a fully automated metagenomics pipeline, from reads to bins. *CNB Centro Natl De Biotecnología* 1–69. <https://doi.org/10.17504/protocols.io/hfqb3mw>
- Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38(7):3022–3027. <https://doi.org/10.1093/molbev/msab120>
- Teoh WK, Salleh FM, Shahir S (2017) Characterization of *Thiomonas delicata* arsenite oxidase expressed in *Escherichia coli*. 3 *Biotech* 7(2):1–7. <https://doi.org/10.1007/s13205-017-0740-7>
- Thi P, Vahlis A (2017) *Comparative metagenomics of PHA synthase genes in soil*
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16):5261–5267. <https://doi.org/10.1128/AEM.00062-07>
- Yang C, Zhang W, Liu R, Zhang C, Gong T, Li Q, Wang S, Song C (2013) Analysis of polyhydroxyalkanoate (PHA) synthase gene and PHA-producing bacteria in activated sludge that produces PHA containing 3-hydroxydodecanoate. *FEMS Microbiol Lett* 346(1):56–64. <https://doi.org/10.1111/1574-6968.12201>
- Yasuda S, Suenaga T, Orschler L, Agrawal S, Lackner S, Terada A (2021) Metagenomic insights into functional and taxonomic compositions of an activated Sludge Microbial Community Treating Leachate of a completed landfill: a pathway-based analysis. *Front Microbiol* 12(April):1–13. <https://doi.org/10.3389/fmicb.2021.640848>
- Ye Y, Doak TG (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 5(8):1–8. <https://doi.org/10.1371/journal.pcbi.1000465>
- Zher Neoh S, Fey Chek M, Tan T, Linares-Pastén H, Nandakumar JA, Hakoshima A, T., Sudesh K (2022) Polyhydroxyalkanoate synthase (PhaC): The key enzyme

for biopolyester synthesis. *Current Research in Biotechnology*, 4(November 2021), 87–101. <https://doi.org/10.1016/j.crbiot.2022.01.002>
Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62(2):316–322. <https://doi.org/10.1128/aem.62.2.316-322.1996>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.