



Developing Bioprospecting Strategies for Bioplastics Through the Large-Scale Mining of Microbial Genomes

Paton Vuong¹, Daniel J. Lim², Daniel V. Murphy¹, Michael J. Wise^{3,4}, Andrew S. Whiteley⁵ and Parwinder Kaur^{1*}

¹ UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, ² Department of Physics and Astronomy, Curtin University, Perth, WA, Australia, ³ School of Physics, Mathematics and Computing, The University of Western Australia, Perth, WA, Australia, ⁴ Marshall Centre for Infectious Disease Research and Training, University of Western Australia, Perth, WA, Australia, ⁵ Centre for Environment and Life Sciences, CSIRO, Floreat, WA, Australia

OPEN ACCESS

Edited by:

Obulisamy Parthiba Karthikeyan,
University of Houston, United States

Reviewed by:

Rajesh K. Sani,
South Dakota School of Mines
and Technology, United States
Vijay Kumar,
Institute of Himalayan Bioresource
Technology (CSIR), India

*Correspondence:

Parwinder Kaur
parwinder.kaur@uwa.edu.au

Specialty section:

This article was submitted to
Microbiotechnology,
a section of the journal
Frontiers in Microbiology

Received: 19 April 2021

Accepted: 21 June 2021

Published: 12 July 2021

Citation:

Vuong P, Lim DJ, Murphy DV,
Wise MJ, Whiteley AS and Kaur P
(2021) Developing Bioprospecting
Strategies for Bioplastics Through
the Large-Scale Mining of Microbial
Genomes.
Front. Microbiol. 12:697309.
doi: 10.3389/fmicb.2021.697309

The accumulation of petroleum-based plastic waste has become a major issue for the environment. A sustainable and biodegradable solution can be found in Polyhydroxyalkanoates (PHAs), a microbially produced biopolymer. An analysis of the global phylogenetic and ecological distribution of potential PHA producing bacteria and archaea was carried out by mining a global genome repository for PHA synthase (PhaC), a key enzyme involved in PHA biosynthesis. Bacteria from the phylum Actinobacteria were found to contain the PhaC Class II genotype which produces medium-chain length PHAs, a physiology until now only found within a few *Pseudomonas* species. Further, several PhaC genotypes were discovered within Thaumarchaeota, an archaeal phylum with poly-extremophiles and the ability to efficiently use CO₂ as a carbon source, a significant ecological group which have thus far been little studied for PHA production. Bacterial and archaeal PhaC genotypes were also observed in high salinity and alkalinity conditions, as well as high-temperature geothermal ecosystems. These genome mining efforts uncovered previously unknown candidate taxa for biopolymer production, as well as microbes from environmental niches with properties that could potentially improve PHA production. This *in silico* study provides valuable insights into unique PHA producing candidates, supporting future bioprospecting efforts toward better targeted and relevant taxa to further enhance the diversity of exploitable PHA production systems.

Keywords: bioplastic, bioprospecting, genome mining, PHA synthase, polyhydroxyalkanoates

INTRODUCTION

Plastic waste pollution has become a significant issue, with plastic waste as of 2015 amounting to 6,300 million metric tons predicted to almost double to 12,000 million metric tons by 2050, with only 9% being recycled, 12% disposed of through incineration and 79% of which is sent to landfill or ending up in the natural environment (Geyer et al., 2017). The disposal processes for petroleum-based plastics are notoriously problematic—incineration is costly and produces toxic by-products

and recycling is slow due to sorting required to accommodate the wide range of plastic formulations with certain plastic additives often limiting their use in recycling (Anjum et al., 2016). In addition, the production of plastics accounts for 4–8% of the global consumption of oil, with the projected use rising to 20% by 2050 (Narancic et al., 2020). The need for biodegradable and sustainable alternatives has, therefore, become critical with microbially produced polyhydroxyalkanoates (PHAs) showing significant promise as an economically viable replacement for petroleum-based plastic (Meereboer et al., 2020).

Polyhydroxyalkanoates (PHAs) are a group of microbially-made polyesters: 100% biodegradable, thermoplastic, insoluble in water, non-toxic and biocompatible. The major advantage for waste management is that PHA products are 100% biodegradable (compostable bioplastics) in the land and ocean, leaving no lasting waste management footprint. As such these biopolymers are well suited as a “green” alternative to petroleum-based plastics by being both biodegradable and non-toxic (Dietrich et al., 2017). Prokaryotic species that produce PHAs are broad and include many bacteria, such as *Alcaligenes latus*, *Ralstonia eutropha*, *Azotobacter beijerinckii*, *Bacillus megaterium* and *Pseudomonas oleovorans* (Bhuwal et al., 2013; Anjum et al., 2016), as well as several archaea from the family *Halobacteriaceae* (Han et al., 2010; Hermann-Krauss et al., 2013). PHAs are produced by microorganisms as a form of intracellular energy storage, within the microbial cytoplasm, and manifests when there is an excess supply of carbon but other essential nutrients, such as oxygen, nitrogen, and phosphorus are deficient (Laycock et al., 2014). The carbon-based “feedstocks” used in the development of efficient microbial PHA production are derived from sustainable and low-cost sources: agricultural waste (starch, lignocellulose and animal carcasses), molasses, whey, waste oils and glycerol from the production of biodiesel (Cui et al., 2016).

The types of PHAs produced by individual microorganisms are categorized by the number of carbon atoms within the monomer units that form the PHA polymer: short-chain length (SCL) PHAs consist of 3–5 carbon atoms, medium-chain length (MCL) PHAs with 6–14 carbon atoms and PHAs with > 14 carbon atoms are considered long-chain length (LCL) (Sagong et al., 2018). The mechanical properties of PHAs are influenced by the carbon length of the monomers, SCL-PHAs due to their crystalline structure, are generally stiff, brittle and have poor thermal stability, requiring modified processing approaches for an improved product (Wang et al., 2016). In comparison, MCL-PHAs have good thermos-elastomeric properties but are mainly limited to bacteria from the family *Pseudomonadaceae* (Oliveira et al., 2020). LCL-PHAs are rare and only produced by *Pseudomonas* strains (Meereboer et al., 2020) and have seen less interest within the area of bioplastic development (Muneer et al., 2020).

The enzyme responsible for the polymerization of monomeric substrates during PHA biosynthesis is PHA synthase (PhaC) which is organized into four classes depending on their primary protein structure, substrate specificity and subunit composition (Rehm, 2003). Class I contains a type of PhaC (~60 kDa) which forms a homodimer; Class II has two synthases PhaC1 and PhaC2 (~60 kDa each), in which only PhaC1 is physiologically active;

Class III forms a heterodimer with a catalytic subunit PhaC (40–53 kDa) and a secondary subunit PhaE (20–40 kDa); and lastly, Class IV also forms a heterodimer with a catalytic subunit PhaC (41.5 kDa) and a secondary subunit PhaR (22 kDa) (Mezzolla et al., 2018). Class I and IV PhaCs favor SCL monomers, with carbon chain lengths C3–C5 whilst Class II PhaC prefer MCL monomers, composed of carbon chain lengths C6–C14 (Chek et al., 2017). Class III PhaCs also favor C3–C5 SCL monomers but can also utilize substrates with C6–C8 carbon chain lengths (Rehm, 2003; Ray and Kalia, 2017).

More than 150 unique monomers have been discovered within PHA polymer samples since the discovery of PHA (Agnew and Pflieger, 2013). PHA polymer properties meet the design needs of nearly all synthetic plastics providing considerable industrial product design and manufacturing flexibility. As an example, PHAs have physical properties similar to synthetic polyethylene and polypropylene (polyolefins) which are used extensively in single-plastic use products which are at higher risk of becoming waste. The composition of the monomers is tied to the substrate specificity of the individual PhaC enzyme, and PHAs with novel monomer compositions are constantly being discovered from bacteria in various environments (Sharma et al., 2017). PhaCs with broad substrate specificity are desirable for two major reasons. PHA with co-polymer composition have been observed to possess better physical properties compared to homopolymer SCL-PHAs (Chek et al., 2017; Ray and Kalia, 2017). In addition, they can more readily utilize a wider range of carbon substrates which could potentially include inexpensive plant or animal waste products (Nielsen et al., 2017; Surendran et al., 2020).

Microbial bioprospecting of large genome collections for the presence of PhaCs could be a viable approach to discover new substrate specificities and monomer compositions that could improve current PHA production systems. Genome mining is a method of *in silico* bioprospecting where collections of genome sequences are searched for putative enzymes involved in the biosynthesis of secondary metabolites (Ziemert et al., 2016). The Integrated Microbial Genomes and Microbiomes system (IMG/M) is an online platform that allows users access to a global repository of genome and metagenome datasets (Chen et al., 2021). The Genomes OnLine Database (GOLD), provided through IMG/M, is a manually curated database with a metadata reporting system that allows users to easily tabulate and browse the associated metadata assigned to each submitted genome (Mukherjee et al., 2021). Exploring the global distribution of microorganisms along with access to large scale ecological data could give valuable insight toward taxa that are less studied, as well as environmental microbes that exist in extreme or unusual ecosystems, for their potential in improving current PHA production methods.

MATERIALS AND METHODS

Preparation of Bacterial and Archaeal Data

A set of 16,576 bacteria (accessed 14th Dec 2020) and 1932 archaea (accessed 20th Feb 2021) genome sequences along

with their associated taxonomic and ecosystem metadata were obtained from the IMG/M¹ warehouse to create the experimental dataset (**Supplementary Data 1, 2**). Due to the large volume of bacterial sequences available, only the JGI sequenced bacteria genomes were obtained for this study, whereas all available archaeal genomes (both JGI and externally sequenced) were downloaded. PhaC protein sequences used to query the genome sequences were sourced from UniprotKB² database using the search term “gene:phac” and included both reviewed (Swiss-Prot) and unreviewed (TrEMBL) results (The UniProt C, 2021). The bacterial PhaC protein sequences were filtered using the term “taxonomy:bacteria” with a resulting 4,038 sequences (accessed 21st Jan 2021) and archaeal sequences with “taxonomy:archaea” resulting in 178 sequences (accessed 20th Feb 2021). Available protein sequences retrieved from UniprotKB were limited to Class I, II, and III PhaCs for bacterial sequences and Class I and III for archaeal sequences. Bash³ and MATLAB⁴ were used to conduct the sorting and categorization of the large-scale data. The scripts are available in our GitHub⁵ repository.

PhaC Reference Sequence Preparation and Curation

The sequences were first manually curated through visual inspection of the metadata. The following categories of sequences were removed through inspection of the protein name: (1) Names determined to not be PhaCs, (2) Names with ambiguous function, (3) Putative PhaCs and (4) Fragment PhaCs. PhaC sequences with amino acid lengths < 328 in bacteria and < 100 in archaea were also filtered out as “fragments.” Duplicate protein sequences were removed. Subsets of the PhaC sequences created according to the class listed in the protein name field from the respective metadata entry. Where the protein name entry did not identify the class, the TIGRFAM metadata entry was used to assign the PhaC class.

BLAST + v2.10.1 was used to classify the remaining unclassified PhaCs (Camacho et al., 2009). The combined subset of known PhaC (Class I, II, and III for bacterial PhaC and Class I and III for archaeal PhaCs), were aligned *via* BlastP with e-value = 1×10^{-10} against the unclassified PhaC sequences of the respective taxonomic domain. The unclassified sequences were then classified according to the class of the query sequence with the lowest e-value. In cases where multiple hits achieved the lowest e-value, the query sequence with the highest bitscore was used for classification. The final curated protein sequence dataset used for the prediction of PhaCs in the genome datasets were 2,799 bacterial PhaCs (Class I = 1,799; Class II = 604; Class III = 375; Unknown = 1) and 140 archaeal PhaCs (Class I = 2; Class III = 137; Unknown = 1). The final curated bacteria and archaea query PhaC protein sequences and associated metadata are available in **Supplementary Data 3, 4**, respectively.

¹<https://img.jgi.doe.gov/cgi-bin/m/main.cgi>

²<https://www.uniprot.org/>

³<https://www.gnu.org/software/bash/>

⁴<https://www.mathworks.com>

⁵https://github.com/pvuonguwa/phac_dist

PhaC Distribution in the Genome Datasets

The PhaC protein sequences were aligned against their respective domain genome sets (**Supplementary Data 1, 2**) using TblastN with e-value = 1×10^{-10} . A single best hit was chosen from each genome as the representative PhaC protein sequence for use in downstream phylogenetic comparisons by selecting the lowest e-value result. In cases where multiple hits achieved the lowest e-value, the hit with the highest bitscore was used as the representative predicted PhaC sequence. The UniprotKB ID of the matching query, as well as PhaC Class, were appended onto the taxonomic and ecosystem metadata of IMG genomes for identified PhaC genotypes and can be found in **Supplementary Data 5, 6** for bacteria and archaea, respectively. Multiple sequence alignment was performed using both the BLAST predicted PhaC sequences within genomes and the curated Uniprot reference sequences *via* Kalign 3, with default settings (Lassmann, 2020). The alignment was performed for all PhaC protein sequences within their respective domains to visualize the distribution of PhaC classes. Due to the large number of sequences, FastTree v2.1.10 (Price et al., 2010) was used to create maximum-likelihood phylogenetic trees using the WAG amino acid substitution model, implemented through the Galaxy⁶ online platform (Afgan et al., 2016). Phylogenetic trees were visualized using the Interactive Tree of Life⁷ (iTOL) online tool (Letunic and Bork, 2019).

RESULTS

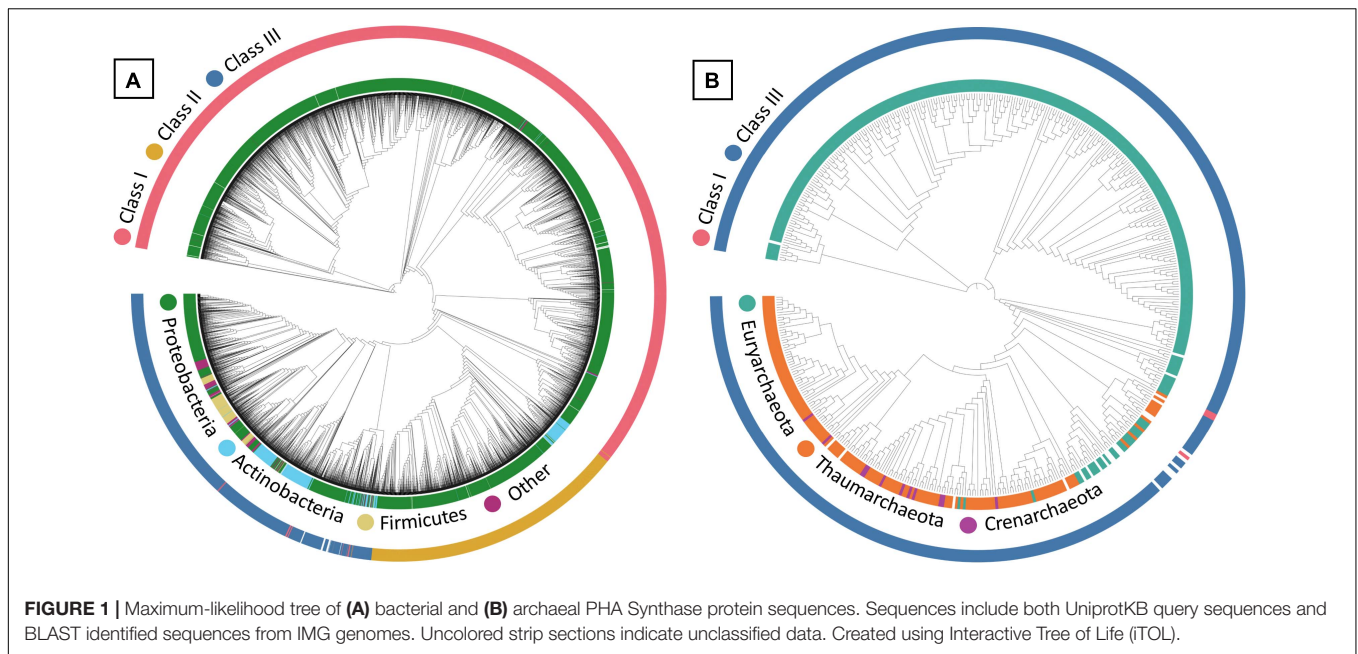
Phylogenetic Distribution of Bacterial PhaCs Genotypes

Within the maximum-likelihood tree, the bacterial PhaC protein sequences appeared to form generally distinct groupings within their respective classes in the phylogenetic tree, with Class III PhaCs showing relatively high variation in phyla compared to those seen in Class I and II PhaCs (**Figure 1A**). PhaCs of unknown class were included to determine if their classification could be inferred *via* phylogenetic tree reconstruction. From the phylogenetic tree, the unknown class sequences appear to fall within the majority Class III grouping on the tree. However, the placements appear to be close to a variable region where different PhaC classes are dispersed and as such tentatively remained unclassified and were omitted from further analysis.

From the initial bacterial genome dataset PhaCs were identified in 24.8% ($n = 4,119$) of the genomes. From this subset of genomes with PhaC genotypes, 2339 Class I, 529 Class II and 1,229 Class III PhaCs were identified according to the class of the matching query sequence used in the BLAST search (**Supplementary Data 5**). Class I PhaCs from the genome dataset were identified predominantly in Proteobacteria comprising 98.2% ($n = 2,298$) of the Class I containing genomes with the remaining 1.8% comprising of Actinobacteria ($n = 26$),

⁶<https://usegalaxy.org.au>

⁷<https://itol.embl.de>



Bacteroidetes ($n = 12$), Candidatus Dadabacteria ($n = 1$), Chloroflexi ($n = 1$) and Spirochetes ($n = 1$). Within Class II PhaC identified genomes, Proteobacteria were also dominant appearing in 81.5% ($n = 431$) of the genomes with the remaining 18.5% from Actinobacteria ($n = 97$) and Bacteroidetes ($n = 1$). Class III PhaCs in comparison were identified in a much larger variety of phyla although Proteobacteria still remaining the largest contributor at 51.8% ($n = 637$), followed by Actinobacteria 27.7% ($n = 341$), Firmicutes 12% ($n = 148$), with the remainder consisting of established phyla 5.4%; Acidobacteria ($n = 5$), Bacteroidetes ($n = 13$), Chloroflexi ($n = 9$), Cyanobacteria ($n = 17$), Elusimicrobia ($n = 3$), Gemmatimonadetes ($n = 2$), Nitrospirae ($n = 3$), Planctomycetes ($n = 3$), Spirochetes ($n = 9$), Verrucomicrobia ($n = 2$) and from candidatus or unknown phyla 3.1%; candidate division NC10 ($n = 1$), Candidatus Blackallbacteria ($n = 3$), Candidatus Dadabacteria ($n = 1$), Candidatus Falkowbacteria ($n = 3$), Candidatus Kaiserbacteria ($n = 2$), Candidatus Melainabacteria ($n = 3$), Candidatus Moranbacteria ($n = 1$), Candidatus Riflebacteria ($n = 1$), Candidatus Rokubacteria ($n = 15$), Candidatus Wallbacteria ($n = 1$), and unclassified ($n = 6$).

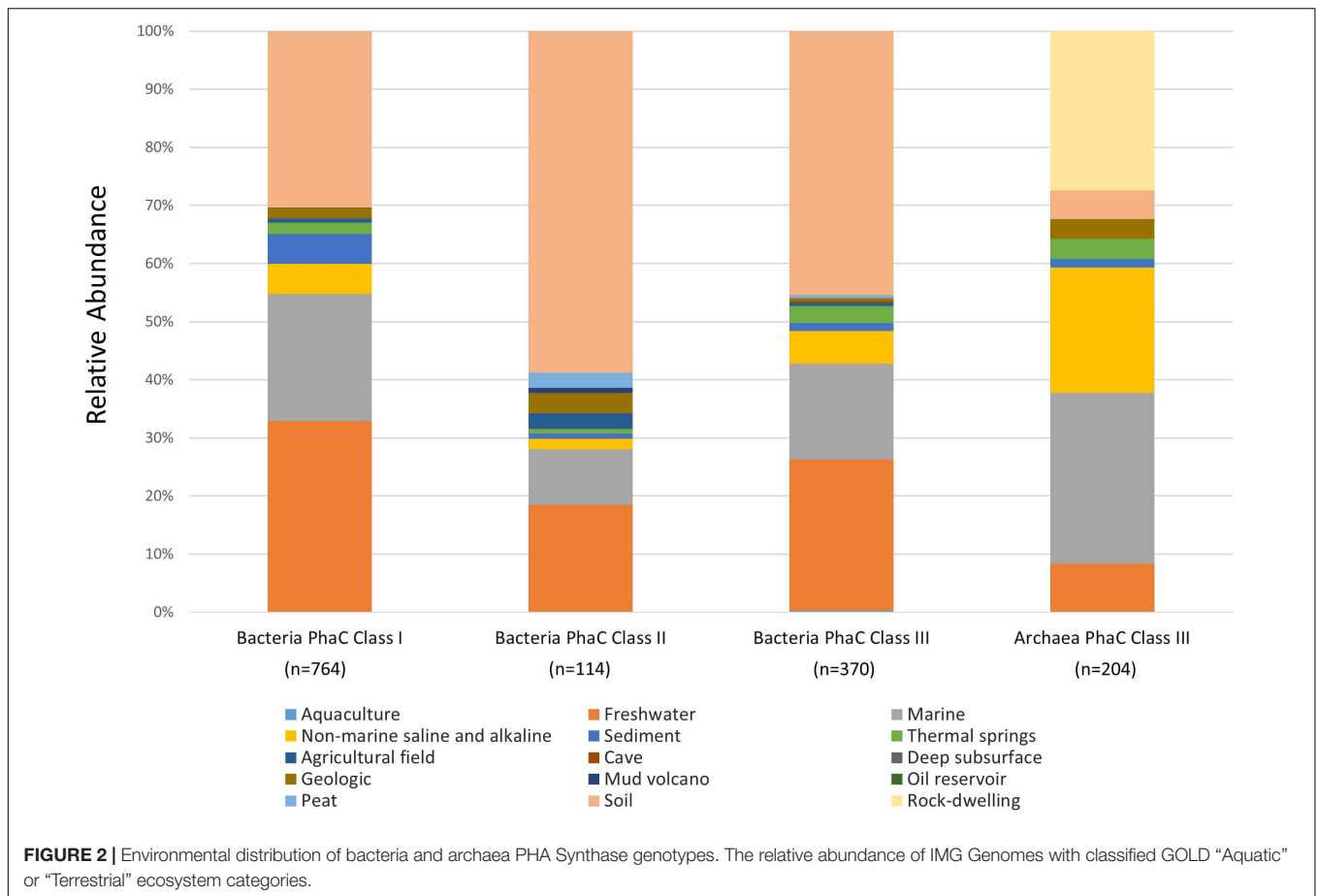
Environmental Distribution of Bacterial PhaCs Genotypes

To determine the distribution of bacteria PhaC genotypes in relevant environments, genomes were used if marked as “environmental,” under the GOLD ecosystem category, or “aquatic” or “terrestrial” under the GOLD ecosystem category, in the metadata contained in **Supplementary Data 5**. Entries that were “unclassified” were not pertinent regarding environmental information and were omitted. From the identified PhaC genotypes, environmental genomes were observed containing 765 Class I, 114 Class II and 370 Class III PhaCs (**Figure 2**).

Class I PhaC genotypes were most abundant in diverse environments with 32.8% ($n = 251$) present in “Freshwater,” 30.4% ($n = 232$) in “Soil” and 21.8% ($n = 167$) from “Marine” ecosystem types. Class I were also present in other aquatic ecosystem types consisting of; 5.2% ($n = 40$) “Non-marine saline and alkaline,” 5.1% ($n = 39$) “Sediment” and 2% ($n = 15$) “Thermal springs.” The remaining 2.7% appeared in small amounts from terrestrial environments; ($n = 5$) “Agricultural field,” ($n = 1$) “Deep subsurface” and ($n = 1$) “Oil reservoir.” Within Class I bacterial genotypes, the presence of Proteobacteria were prevalent across all observed ecosystem types, with trace representations of Actinobacteria in “Soil” and “Thermal springs,” Bacteroides in “Freshwater” and “Soil,” Chloroflexi in “Marine,” Spirochetes in “Freshwater” and Candidatus Dadabacteria in “Soil” (**Supplementary Table 1**).

Over half of Class II PhaC genotypes were found in “Soil,” with 58.8% ($n = 67$) representation with the second and third largest percentage consisting of diverse aquatic environments, 18.4% ($n = 21$) “Freshwater” and 9.6% ($n = 11$) “Marine” ecosystem types. The rest of the Class II genotypes were found in the remaining 7% terrestrial environments; ($n = 4$) “Geologic,” ($n = 3$) “Agricultural field,” ($n = 3$) “Peat” and ($n = 1$) “Mud volcano” as well as 3.6% aquatic environments; ($n = 2$) “Non-marine saline and alkaline,” ($n = 1$) “Sediment” and ($n = 1$) “Thermal springs.” From the Class II genotypes, Proteobacteria was present across almost all observed ecosystem types, with Actinobacteria being the only other phylum present, found mostly within “Soil” and in trace amount across “Freshwater,” “Marine” and “Geologic” and the sole representative within “Mud volcano” (**Supplementary Table 2**).

Class III PhaC genotypes followed the trend of having the highest percentage of abundance in diverse terrestrial and aquatic environments with 45.4% ($n = 168$) in “Soil,” 25.9% ($n = 96$) “Freshwater” and 16.6% ($n = 61$) from “Marine” ecosystem types.



Class III genotypes were present in other aquatic environments with relative abundances of; 5.7% ($n = 21$) “Non-marine saline and alkaline,” 3% ($n = 11$) “Thermal springs,” 1.4% ($n = 5$) “Sediment” and 0.3% ($n = 1$) “Aquaculture” with the remaining 1.8% from terrestrial ecosystem types; ($n = 2$) “Agricultural field,” ($n = 2$) “Geologic” and ($n = 1$) “Cave.” Compared to Class I and II, Class III genotypes were highly diverse across the observed ecosystem types (Supplementary Table 3). The ecosystem types containing Class III genotypes of note were “Freshwater” and “Soil,” with the highest levels of diversity across all PhaC classes and were observed to contain representatives of several Candidatus phyla.

Environmental bacterial PhaC genotypes also investigated to determine whether they were recovered from unusual or niche settings (Table 1). Within each class a total of 8.6% ($n = 66$) Class I, 5.4% ($n = 6$) Class II and 10.8% ($n = 40$) Class III PhaC genotypes were found in ecosystems with extreme physicochemical conditions. The most common niche environments present across all bacterial PhaC genotypes were high salinity and alkalinity conditions found in “Non-marine saline and alkaline” ecosystem types and the unique high-temperature geothermal environments found in “Thermal springs” and “Mud volcano” ecosystem types as well as “Hydrothermal vents” found in “Marine” ecosystem types. Bacteria PhaC genotypes were also found

in acidic, desert, saline or permafrost environments, as well as ecosystems with petroleum contamination: “Oil reservoir” and “Creosote-contaminated soil.” The distribution of phyla across these niche ecosystem settings for Class I, II and III bacterial genotypes can be found in Supplementary Tables 1–3, respectively.

Phylogenetic Distribution of Archaeal PhaCs Genotypes

In the maximum-likelihood tree, most of the archaeal PhaC protein sequences belonged to Class III PhaCs, with the few Class I and unclassified Class PhaCs sequences appeared to group closely together apart from the Class III PhaCs (Figure 1B). Due to the ambiguity of the unknown class PhaCs, they were omitted from further downstream analysis. Within the Class III PhaCs, a distinct grouping can be observed from Class III PhaCs belonging to Euryarchaeota compared to the Class III PhaCs found in Crenarchaeota and Thaumarchaeota. From the initial dataset of archaeal genomes, there were PhaCs predicted in 17.5% ($n = 338$) of the samples (Supplementary Data 6). The sole genome containing Class I ($n = 1$) belonged to the phylum Euryarchaeota. Class III was the dominant PhaC genotype appearing within the archaea genomes with 65.4% belonging to Euryarchaeota ($n = 217$), 30.1% from Thaumarchaeota ($n = 100$), 3% from

Crenarchaeota ($n = 10$) and the remaining 1.5% unclassified ($n = 5$).

Environmental Distribution of Archaeal PhaCs

Within the archaeal PhaC identified genomes, only Class III PhaC genotypes had representative samples from environmental settings according to the GOLD metadata listed in **Supplementary Data 6 (Figure 2)**. The highest abundance environments that archaea Class III genotypes were present in were 29.4% ($n = 60$) “Marine,” 27.4% ($n = 56$) “Rock-dwelling” and 21.6% ($n = 44$) “Non-marine saline and alkaline” ecosystem types. Contrary to the bacteria PhaCs, archaea Class III PhaCs

genotypes were less represented in the diverse ecosystem with abundances of 8.3% ($n = 17$) in “Freshwater” and 4.9% ($n = 10$) from “Soil” with the remaining found in 3.4% ($n = 7$) “Thermal springs,” 3.4% ($n = 7$) “Geologic” and 1.6% ($n = 3$) from “Sediment” ecosystem types. Euryarchaeota were present across all observed ecosystem types, with Thaumarchaeota present in “Freshwater,” “Marine,” “Thermal springs” and “Soil,” Crenarchaeota present solely in “Marine,” and unclassified archaeal phyla within “Marine” and “None-marine saline and alkaline” (**Supplementary Table 4**).

Over half of the archaea Class III genotypes were found in extreme settings with 57.8% of all archaea PhaC environmental genomes from ecosystems with extreme physicochemical conditions (**Table 1**). The two settings that formed the majority of extreme ecosystems were the saline “Halite pinnacle” found in almost all the “Rock-dwelling” ecosystem types as well as the high salinity and alkalinity “Non-marine saline and alkaline” ecosystem types. Aside from other saline environments, archaea PhaC genotypes were also found in the “Marine” ecosystems within “deep sea,” and “Hydrothermal vents” habitats as well as from “Thermal springs” ecosystems. The distribution of archaeal phyla within the observed niche settings can be found in **Supplementary Table 4**.

TABLE 1 | Environmental PhaC genotypes found in extreme physicochemical conditions.

Ecosystem Type	Extreme ecosystem subtype or habitat descriptors	Bacteria PhaC			Archaea PhaC
		Class I	Class II	Class III	Class III
Marine	Deep oceanic, basalt-hosted subsurface hydrothermal fluid, Hydrothermal vents	3	2	4	3
	Deep-sea Creosote-contaminated soil			1	3
Non-marine saline and alkaline		40	2	21	44
Thermal springs		15	1	11	7
Geologic	Acid Mine	2			
	Saline water, Salt, and Salt mine	4			3
Oil reservoir		1			
Mud volcano			1		
Rock-dwelling	Halite pinnacle				55
Soil	Creosote-contaminated soil			1	
	Desert soil	1		1	
	Permafrost sediment			1	
	Saline				3
Subtotal (Relative abundance)		66 (8.6%)	6 (5.3%)	40 (10.8%)	118 (57.8%)

Descriptors are taken from GOLD metadata under the ecosystem subtype or habitat headings. Ecosystems with no descriptors inherently contain extreme conditions.

DISCUSSION

The genome mining effort has provided an encompassing view of the phylogenetic and ecological distribution of bacterial and archaeal PhaC genotypes. This *in silico* microbial bioprospecting method presents a broad categorization method that inexpensively searches publically available data for PHA producing candidates across multiple bacterial and archaeal phyla. Through this approach, we were able to visualize the presence of different PhaC classes amongst various phyla and ecosystems. The resulting data provided insights on the general distribution of PhaC genotypes including ecological “hot spots,” as well as presenting microbial candidates that could improve current PHA production systems by utilizing cheaper feedstock or providing more efficient operating conditions.

PhaC genotypes appeared in a broad number of bacterial and archaeal genomes, consisting of 24.8 and 17.5%, respectively of the genome datasets, suggesting that PHAs are a relatively conserved form of carbon energy storage within bacteria (Serafim et al., 2018) and archaea (Wang et al., 2019b). This is also supported by the distribution of Class III PhaCs in bacteria, which the study identified across 23 phyla, roughly one-fifth of the 111 recognized bacterial phyla by the Genome Taxonomy Database⁸ through the standardization of bacterial taxonomy (Parks et al., 2018). Class III PhaCs have more variation in PHA chain lengths compared to other SCL-PHA producing classes, which may lead to a great potential of unique monomer subunits (Ray and Kalia, 2017). Combined with the large scale of observed diversity, Class III PhaCs may harbor a wider range of substrate specificities that could potentially produce novel monomeric compositions which

⁸<https://gtdb.ecogenomic.org/>

could be the target for protein engineering for improved PHA properties and production (Jia et al., 2016). The findings suggest that PhaCs, in particular Class III genotypes, are more prevalent and diverse in composition within both bacteria and archaea domains through our *in silico* observations.

The dominance of Proteobacteria across Class I and II and roughly half of Class III PhaC genotypes suggest that PHA production is highly conserved within the phylum. The prevalence of Proteobacteria as PHA producers was similarly observed in a study investigating PHA accumulating organisms in a mixed microbial culture, with Proteobacteria accounting for nearly 88% of the microbial community when the PHA content was at its peak level (Sruamsiri et al., 2020). Proteobacteria have been observed to have highly diversified genomes that enables metabolic adaptability in response to varying environmental conditions (Zhou et al., 2020). This indicates that Proteobacteria contain innate physiology which more readily responds to changes to external stressors, with PHA production as one of the responses to fluctuations in available nutrient levels. Proteobacteria genotypes contain Class I, II and III PhaCs, which demonstrates a wide variation in substrate specificity, likely owing to the adaptive nature of the diversified genomes. However, this does not explain the almost exclusivity of Class I and II PhaCs to Proteobacteria and the relatively high taxonomic diversity seen in Class III PhaCs. Within the PHA gene cluster, *phaC* has been identified as the hub gene and plays a major role in cluster formation in prokaryotes (Kutralam-Muniasamy et al., 2017). Further study is required to uncover why Class III PhaC genotypes presumably undergoes more horizontal transfer compared to Class I and II and why Class III cluster appeared to be favored in a higher variety of taxa.

A potential MCL-PHA producing alternative the study discovered is bacteria from the phylum Actinobacteria, which was amongst the Class II PhaC genotypes that were identified in our studies. Class II PhaCs are of particular interest as they are the only PhaC class capable of producing MCL-PHAs and bacteria that generate MCL-PHAs are mainly *Pseudomonas* which consists of almost all of the Proteobacteria PhaC II genotypes that were found in this study (Mozejko-Ciesielska et al., 2019). MCL-PHAs have not had widespread production due to requiring expensive carbon sources needed for *Pseudomonas* PHA production, meaning cheaper compatible feedstocks were needed to be found (Tanikkul et al., 2020), or recombinant strains are required to utilize inexpensive industry by-products as substrates (Löwe et al., 2017). Actinobacteria is widely recognized within sustainable industries for its ability to breakdown plant biomass including the use of agricultural and forestry plant waste as a carbon source (Lewin et al., 2016; Bettache et al., 2018). Since raw materials can account for 40–48% of production costs, bioprospecting for Actinobacteria Class II PhaC genotypes could potentially kickstart industrial levels of MCL-PHAs through the use of inexpensive feedstock (Schmidt et al., 2016).

Roughly two-thirds of the archaea PhaC genotypes found in this study were from the phylum Euryarchaeota, of which most of the constituents were class Halobacteria. The abundance of Halobacteria PhaC genotypes and the favourability of extremophiles to work in broad conditions is reflected in

the almost exclusive focus on haloarchaea in the study of archaea PHA producers (Han et al., 2010; Legat et al., 2010; Hermann-Krauss et al., 2013). Our search identified several PhaC genotypes within the archaeal phylum Thaumarchaeota within aquatic ecosystems, and to date, no extensive studies have been performed on the viability of PHA production within this phylum. Thaumarchaeota are found across many ecosystems, including extreme environments such as acidic and alkaline soils as well as in trenches within the hadal zone of the ocean (Wang et al., 2019a; Zhong et al., 2020). The CO₂ fixation pathway is highly efficient even in nutrient-limited conditions within Thaumarchaeota, meaning it could potentially utilize CO₂ as a cheap and sustainable feedstock for PHA production (Könneke et al., 2014; Reji and Francis, 2020). With a broad range of operation and efficient utilization of an inexpensive carbon source, Thaumarchaeota could be considered a viable PHA production candidate.

The environment with extreme conditions that PhaC genotypes from both bacteria and archaea were observed to appear most in was the “non-marine saline and alkaline” ecosystem type. Soda lakes are the prime example of this ecosystem type being saline, high pH environments with high amounts of available carbon in the form of carbonates and bicarbonates, often with limited levels of nitrogen compounds—conditions that favor the production of PHAs (Ogato et al., 2015). Despite appearing as inhospitable environments, they are host to a diverse array of microbial life, which includes multiple bacterial phyla and the species from the Euryarchaeota phylum of archaea (Sorokin et al., 2014). With the right conditions for PHA production as well as positive genotyping of PhaCs from both domains from our study, high salinity, high pH environments such as soda lakes could be a potential trove of PHA producing microbes for future microbial bioprospecting ventures.

Another extreme physicochemical environment of note from our findings into the ecological distribution of PhaC genotypes were geothermal ecosystems such as hydrothermal vents or hot springs. Hydrothermal vents are areas of the ocean floor where seawater percolates through cracks in the crust and are superheated by rocks in contact with magma (Dick, 2019). The areas, around which hydrothermal fluid circulate, are rich in carbon and are dominated by microbial thermophilic methanogens (Minic and Thongbam, 2011; Ver Eecke et al., 2012). Hot springs are similar in setting, except the water source comes from groundwater and is also rich in microbial life (Urschel et al., 2015; Narsing Rao et al., 2021). From observation of environmental PhaC genotypes from niche environments, it seems that PHA producers can be readily found in extremophiles containing extreme physicochemical conditions, so long as there is an abundant carbon source.

Over half of the archaea PhaC genotypes were found in extreme environments, which may indicate that taxa that commonly contain extremophiles may have developed physiological strategies utilizing PHA production. Within extreme environments, PHA production in microbial extremophiles acts not only as a carbon energy reserve but may also provide a protective role by maintaining cellular and membrane integrity against extreme physicochemical

conditions (Obulisamy and Mehariya, 2021). PHA has been shown to function alongside extracellular polysaccharides (EPS) in microbial extremophiles to protect cellular function and structure from physiochemical stress (López-Ortega et al., 2021). Studies have found that in *Haloferax mediterranei*, a species of halophilic archaea, changing stressors such as increasing salinity (Cui et al., 2017a) or making nitrogen levels deficient (Cui et al., 2017b) can promote the accumulation of PHA over that of EPS. These findings provide insight into how PHA performs a critical role in microbial survival against external stress factors, as well as strategies on how to apply these stress factors to improve accumulation rate during PHA production.

PhaC genotypes discovered from these extreme environments could prove beneficial for cost reduction in industrial PHA production. Factors that contribute to the high cost of PHA production include maintaining a closed, contamination-free environment for fermentation/growth, expensive feedstock and difficult PHA extraction processes (Chen et al., 2020). Utilizing halophile PHA producers conveys several advantages in this regard: a high salinity environment can undergo continuous fermentation in an open unsterile environment which reduces operating costs, they can use various inexpensive feedstocks to produce PHAs and due to high intracellular osmotic pressure of halophiles, they can be lysed using tap/municipal water for easy PHA extraction (Mitra et al., 2020). Thermophiles can similarly reduce contamination risks by using elevated temperatures during fermentation (Xiao et al., 2015) and have been investigated for use in sustainable industries for their thermostable bioproducts (Debnath et al., 2019). By exploiting the extreme conditions that certain microbes have adapted to, we may be able to develop processes that can lead to more efficient and cost-effective PHA production techniques.

Studies screening for PHA synthase sequences recovered from niche and under-represented environments has yielded success in novel insights toward diversifying PHA production methods. Genomes from *Janthinobacterium* sp. isolates recovered from Antarctica were found to contain hitherto uncharacterized putative PHA synthase genes that were phylogenetically divergent from the known PhaC classes and tentatively labeled as “Class V” PhaCs (Tan et al., 2020). The first occurrence of PHA producers from hypersaline microbial mats were discovered in salt concentration ponds located in Mexico through PCR amplification of *phaC* genes (Martínez-Gutiérrez et al., 2018). A study which mined PHA synthases from mangrove soil metagenomes in Malaysia uncovered a novel synthase from unculturable bacteria that was able to utilize a wide range of substrates by incorporating six types of PHA monomers, demonstrating that PHA sequences procured from *in silico* based approaches are effective for diversifying PHA production methods (Foong et al., 2018). The aforementioned studies demonstrate the versatility of *in silico* screening methods for directing future strategies toward selecting promising candidates and novel synthases for PHA production strategies.

The creation of a categorized genomic list of potential PHA producing microbes presents a valuable resource that directs researchers to explicit candidates, as well as access to potential metabolic pathways of interest. A similar bioprospecting strategy

was utilized successfully by Bustamante et al. (2019) using *in silico* screening methods to screen for putative PHA and β -galactosidase genes in a select number of microorganisms from literature to find PHA producers that could hydrolyze lactose for utilizing excess whey produced from the dairy industry as a cheap feedstock alternative. In comparison, our study has presented a broader scope by using genomes from large-scale data repositories in the public domain, allowing researchers access to a wider array of candidates which can be sorted by PhaC class, taxonomy and ecological information. Through this resource, users can filter for candidates of interest for and can locate the resulting genome for further analysis of metabolic activity relevant to the production method being investigated.

This study is a proof of concept that contextual information generated from large-scale genome mining is highly valuable for use in developing bioprospecting strategies with regards to both ecological and taxonomic data. This approach can be applied to any target functional protein or gene of interest to create a shortlist of microbial candidates and their potential ecosystems for further in-depth analysis or experimentation. The effectiveness of this approach is dependent on not only the volume of the available sequence data, but also on prudent reporting and ready access to metadata. Accompanying information such as physiochemical conditions, geolocation and other ecologically contextual data are often missing from sequence data recovered from environments, which is critical for the eventual recovery of isolates (Gutleben et al., 2018). To address this issue, reporting protocols have been proposed such as the Minimum Information about any (x) Sequence (MIxS) by the Genome Standards Consortium⁹ and the GOLD metadata system by utilized by IMG/M being examples of creating standards for robust metadata reporting. The continual growth of online sequence repositories within the public domain and improvement of standards for metadata reporting presents a massive pool of data freely available to exploit. The mining of large-scale data provides a broader view into patterns in environmental or taxonomic distribution of biosynthetic products, which only improves over time as more data is constantly accumulated from studies conducted across the globe.

CONCLUSION

Genome mining allows for the large-scale discovery of potential microbial PHA producers through *in silico* identification of PhaC genotypes. Through the categorization of genotypes *via* phylogenetic and ecological information, we can find insight through the discovery of less-studied taxa as well as microbes from environmental niches with potential properties that can improve on current PHA production techniques. Although this approach is no substitute for protein expression assays, it presents an overview of the trends inherent in PHA producing microbes. The information gained from this study could be utilized for directing bioprospecting ventures for more effective discovery of potential PHA producing microbes.

⁹<https://gensc.org/mixs/>

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PV and DL performed the research under the guidance of PK, DM, MW, and AW. PV wrote the manuscript with contributions from DL, PK, DM, MW, and AW. All authors read the manuscript and approved the content.

FUNDING

We gratefully acknowledge the resources provided by University of Western Australia and additional computational resources and support from the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.697309/full#supplementary-material>

REFERENCES

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Èech, M., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44, W3–W10. doi: 10.1093/nar/gkw343
- Agnew, D. E., and Pflieger, B. F. (2013). Synthetic biology strategies for synthesizing polyhydroxyalkanoates from unrelated carbon sources. *Chem. Eng. Sci.* 103, 58–67. doi: 10.1016/j.ces.2012.12.023
- Anjum, A., Zuber, M., Zia, K. M., Noreen, A., Anjum, M. N., and Tabasum, S. (2016). Microbial production of polyhydroxyalkanoates (PHAs) and its copolymers: a review of recent advancements. *Int. J. Biol. Macromol.* 89, 161–174. doi: 10.1016/j.ijbiomac.2016.04.069
- Bettache, A., Zahra, A., Boucherba, N., Bouiche, C., Hamma, S., Maibeche, R., et al. (2018). Lignocellulosic biomass and cellulolytic enzymes of actinobacteria. *SAJ Biotechnology* 5:203.
- Bhuwal, A. K., Singh, G., Aggarwal, N. K., Goyal, V., and Yadav, A. (2013). Isolation and screening of polyhydroxyalkanoates producing bacteria from pulp, paper, and cardboard industry wastes. *Int. J. Biomater.* 2013, 752821. doi: 10.1155/2013/752821
- Bustamante, D., Segarra, S., Tortajada, M., Ramón, D., Del Cerro, C., Auxiliadora Prieto, M., et al. (2019). In silico prospection of microorganisms to produce polyhydroxyalkanoate from whey: *caulobacter segnis* DSM 29236 as a suitable industrial strain. *Microb. Biotechnol.* 12, 487–501. doi: 10.1111/1751-7915.13371
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chek, M. F., Kim, S.-Y., Mori, T., Arsad, H., Samian, M. R., Sudesh, K., et al. (2017). Structure of polyhydroxyalkanoate (PHA) synthase PhaC from *Chromobacterium* sp. USM2, producing biodegradable plastics. *Sci. Rep.* 7:5312. doi: 10.1038/s41598-017-05509-4
- Chen, G.-Q., Chen, X.-Y., Wu, F.-Q., and Chen, J.-C. (2020). Polyhydroxyalkanoates (PHA) toward cost competitiveness and functionality. *Adv. Ind. Eng. Polym. Res.* 3, 1–7. doi: 10.1016/j.aiepr.2019.11.001
- Chen, I. M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi: 10.1093/nar/gkaa939
- Cui, Y.-W., Gong, X.-Y., Shi, Y.-P., and Wang, Z. (2017a). Salinity effect on production of PHA and EPS by *Haloferax mediterranei*. *RSC Advances* 7, 53587–53595. doi: 10.1039/C7RA09652F
- Cui, Y.-W., Shi, Y.-P., and Gong, X.-Y. (2017b). Effects of C/N in the substrate on the simultaneous production of polyhydroxyalkanoates and extracellular polymeric substances by *Haloferax mediterranei* via kinetic model analysis. *RSC Advances* 7, 18953–18961. doi: 10.1039/C7RA02131C
- Cui, Y.-W., Zhang, H.-Y., Lu, P.-F., and Peng, Y.-Z. (2016). Effects of carbon sources on the enrichment of halophilic polyhydroxyalkanoate-storing mixed microbial culture in an aerobic dynamic feeding process. *Sci. Rep.* 6, 30766–30766. doi: 10.1038/srep30766
- Debnath, T., Kujur, R. R. A., Mitra, R., and Das, S. K. (2019). “Diversity of microbes in hot springs and their sustainable use,” in *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications: Volume 1. Microbial Diversity in Normal & Extreme Environments*, eds T. Satyanarayana, B. N. Johri, and S. K. Das (Singapore: Springer Singapore), 159–186.
- Dick, G. J. (2019). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* 17, 271–283. doi: 10.1038/s41579-019-0160-2
- Dietrich, K., Dumont, M.-J., Del Rio, L. F., and Orsat, V. (2017). Producing PHAs in the bioeconomy — Towards a sustainable bioplastic. *Sust. Prod. Consump.* 9, 58–70. doi: 10.1016/j.spc.2016.09.001
- Foong, C. P., Lakshmanan, M., Abe, H., Taylor, T., Foong, S. Y., and Sudesh, K. (2018). A novel and wide substrate specific polyhydroxyalkanoate (PHA)
- Supplementary Table 1** | Bacteria class I PhaC genotype environmental distribution.
- Supplementary Table 2** | Bacteria class II PhaC genotype environmental distribution.
- Supplementary Table 3** | Bacteria class III PhaC genotype environmental distribution.
- Supplementary Table 4** | Archaea class III PhaC genotype environmental distribution.
- Supplementary Data 1** | List of Bacterial Genomes retrieved from IMG/M with ID and accompanying metadata.
- Supplementary Data 2** | List of Archaeal Genomes retrieved from IMG/M with ID and accompanying metadata.
- Supplementary Data 3** | Curated Uniprot bacteria PhaC protein sequences. Contains protein sequences along with taxonomic metadata and Uniprot ID. Sequences that did not contain PhaC class information from Uniprot were classified via BLAST against PhaCs with known class.
- Supplementary Data 4** | Curated Uniprot archaea PhaC protein sequences. Contains protein sequences along with taxonomic metadata and Uniprot ID. Sequences that did not contain PhaC class information from Uniprot were classified via BLAST against PhaCs with known class.
- Supplementary Data 5** | Dataset containing IMG/M bacterial genomes with PhaC genotypes predicted from BLAST. Each IMG genome contains taxonomic and ecological GOLD metadata as well as the corresponding Uniprot ID from BLAST best hit.
- Supplementary Data 6** | Dataset containing IMG/M archaea genomes with PhaC genotypes predicted from BLAST. Each IMG genome contains taxonomic and ecological GOLD metadata as well as the corresponding Uniprot ID from BLAST best hit.

- synthase from unculturable bacteria found in mangrove soil. *J. Polym. Res.* 25, 1–9. doi: 10.1007/s10965-017-1403-4
- Geyer, R., Jambeck, J. R., and Law, K. L. (2017). Production, use, and fate of all plastics ever made. *Sci. Adv.* 3:e1700782. doi: 10.1126/sciadv.1700782
- Gutleben, J., De Mares, M., van Elsas, J. D., Smidt, H., Overmann, J., and Sipkema, D. (2018). The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.* 44, 212–229. doi: 10.1080/1040841X.2017.1332003
- Han, J., Hou, J., Liu, H., Cai, S., Feng, B., Zhou, J., et al. (2010). Wide distribution among halophilic archaea of a novel polyhydroxyalkanoate synthase subtype with homology to bacterial type III synthases. *Appl. Environ. Microbiol.* 76:7811. doi: 10.1128/AEM.01117-10
- Hermann-Krauss, C., Koller, M., Muhr, A., Fasl, H., Stelzer, F., and Braune, G. (2013). Archaeal production of polyhydroxyalkanoate (PHA) co-and terpolyesters from biodiesel industry-derived by-products. *Archaea* 2013:129268. doi: 10.1155/2013/129268
- Jia, K., Cao, R., Hua, D. H., and Li, P. (2016). Study of class i and class iii polyhydroxyalkanoate (PHA) synthases with substrates containing a modified side chain. *Biomacromolecules* 17, 1477–1485. doi: 10.1021/acs.biomac.6b00082
- Könneke, M., Schubert, D. M., Brown, P. C., Hügl, M., Standfest, S., Schwander, T., et al. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *Proc. Nat. Acad. Sci. U.S.A.* 111:8239. doi: 10.1073/pnas.1402028111
- Kutralam-Muniasamy, G., Corona-Hernandez, J., Narayanasamy, R.-K., Marsch, R., and Pérez-Guevara, F. (2017). Phylogenetic diversification and developmental implications of poly-(R)-3-hydroxyalkanoate gene cluster assembly in prokaryotes. *FEMS Microbiol. Lett.* 364, 1–9. doi: 10.1093/femsle/fnx135
- Lassmann, T. (2020). Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics* 36, 1928–1929. doi: 10.1093/bioinformatics/btz795
- Laycock, B., Halley, P., Pratt, S., Werker, A., and Lant, P. (2014). The chemomechanical properties of microbial polyhydroxyalkanoates. *Prog. Polym. Sci.* 39, 397–442. doi: 10.1016/j.progpolymsci.2013.06.008
- Legat, A., Gruber, C., Zangger, K., Wanner, G., and Stan-Lotter, H. (2010). Identification of polyhydroxyalkanoates in Halococcus and other haloarchaeal species. *Appl. Microbiol. Biotechnol.* 87, 1119–1127. doi: 10.1007/s00253-010-2611-6
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., et al. (2016). Evolution and ecology of actinobacteria and their bioenergy applications. *Ann. Rev. Microbiol.* 70, 235–254. doi: 10.1146/annurev-micro-102215-095748
- López-Ortega, M. A., Chavarría-Hernández, N., López-Cuellar, M. D. R., and Rodríguez-Hernández, A. I. (2021). A review of extracellular polysaccharides from extreme niches: an emerging natural source for the biotechnology. From the adverse to diverse! *Int. J. Biol. Macromol.* 177, 559–577. doi: 10.1016/j.ijbiomac.2021.02.101
- Löwe, H., Schmauder, L., Hobmeier, K., Kremling, A., and Pflüger-Grau, K. (2017). Metabolic engineering to expand the substrate spectrum of *Pseudomonas putida* toward sucrose. *Microbiologyopen* 6:e00473. doi: 10.1002/mbo3.473
- Martínez-Gutiérrez, C. A., Latisnere-Barragán, H., García-Maldonado, J. Q., and López-Cortés, A. (2018). Screening of polyhydroxyalkanoate-producing bacteria and PhaC-encoding genes in two hypersaline microbial mats from Guerrero Negro. *PeerJ* 6, e4780–e4780. doi: 10.7717/peerj.4780
- Meereboer, K. W., Misra, M., and Mohanty, A. K. (2020). Review of recent advances in the biodegradability of polyhydroxyalkanoate (PHA) bioplastics and their composites. *Green Chem.* 22, 5519–5558. doi: 10.1039/D0GC01647K
- Mezzolla, V., D'Urso, O. F., and Poltronieri, P. (2018). Role of PhaC Type I and Type II Enzymes during PHA Biosynthesis. *Polymers* 10:910. doi: 10.3390/polym10080910
- Minic, Z., and Thongbam, P. D. (2011). The biological deep sea hydrothermal vent as a model to study carbon dioxide capturing enzymes. *Mar. Drugs* 9, 719–738. doi: 10.3390/md9050719
- Mitra, R., Xu, T., Xiang, H., and Han, J. (2020). Current developments on polyhydroxyalkanoates synthesis by using halophiles as a promising cell factory. *Microb. Cell Fact.* 19:86. doi: 10.1186/s12934-020-01342-z
- Mozejko-Ciesielska, J., Szacherska, K., and Marciniak, P. (2019). *Pseudomonas* species as producers of eco-friendly polyhydroxyalkanoates. *J. Polym. Environ.* 27, 1151–1166. doi: 10.1007/s10924-019-01422-1
- Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., et al. (2021). Genomes online database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* 49, D723–D733. doi: 10.1093/nar/gkaa983
- Muneer, F., Rasul, I., Azeem, F., Siddique, M. H., Zubair, M., and Nadeem, H. (2020). Microbial polyhydroxyalkanoates (PHAs): efficient replacement of synthetic polymers. *J. Polym. Environ.* 28, 2301–2323. doi: 10.1007/s10924-020-01772-1
- Narancic, T., Cerrone, F., Beagan, N., and O'Connor, K. E. (2020). Recent advances in bioplastics: application and biodegradation. *Polymers* 12:920. doi: 10.3390/polym12040920
- Narsing Rao, M. P., Dong, Z.-Y., Luo, Z.-H., Li, M.-M., Liu, B.-B., Guo, S.-X., et al. (2021). Physicochemical and microbial diversity analyses of indian hot springs. *Front. Microbiol.* 12:627200. doi: 10.3389/fmicb.2021.627200
- Nielsen, C., Rahman, A., Rehman, A. U., Walsh, M. K., and Miller, C. D. (2017). Food waste conversion to microbial polyhydroxyalkanoates. *Microb. Biotechnol.* 10, 1338–1352. doi: 10.1111/1751-7915.12776
- Obulisamy, P. K., and Mehariya, S. (2021). Polyhydroxyalkanoates from extremophiles: a review. *Bioresour. Technol.* 325:124653. doi: 10.1016/j.biortech.2020.124653
- Ogato, T., Kifle, D., and Lemma, B. (2015). Underwater light climate, thermal and chemical characteristics of the tropical soda lake Chitu, Ethiopia: spatio-temporal variations. *Limnologia* 52, 1–10. doi: 10.1016/j.limno.2015.02.003
- Oliveira, G. H. D., Zaiat, M., Rodrigues, J. A. D., Ramsay, J. A., and Ramsay, B. A. (2020). Towards the production of mcl-PHA with enriched dominant monomer content: process development for the sugarcane biorefinery context. *J. Polym. Environ.* 28, 844–853. doi: 10.1007/s10924-019-01637-2
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Ray, S., and Kalia, V. C. (2017). Microbial cometabolism and polyhydroxyalkanoate co-polymers. *Indian J. Microbiol.* 57, 39–47. doi: 10.1007/s12088-016-0622-4
- Rehm, B. H. A. (2003). Polyester synthases: natural catalysts for plastics. *Biochem. J.* 376(Pt 1), 15–33. doi: 10.1042/BJ20031254
- Reji, L., and Francis, C. A. (2020). Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J.* 14, 2105–2115. doi: 10.1038/s41396-020-0675-6
- Sagong, H.-Y., Son, H. F., Choi, S. Y., Lee, S. Y., and Kim, K.-J. (2018). Structural insights into polyhydroxyalkanoates biosynthesis. *Trends Biochem. Sci.* 43, 790–805. doi: 10.1016/j.tibs.2018.08.005
- Schmidt, M., Ienczak, J. L., Quines, L. K., Zanfonato, K., Schmidell, W., and de Aragão, G. M. F. (2016). Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production in a system with external cell recycle and limited nitrogen feeding during the production phase. *Biochem. Eng. J.* 112, 130–135.
- Serafim, L. S., Xavier, A. M. R. B., and Lemos, P. C. (2018). “Storage of hydrophobic polymers in bacteria,” in *Biogenesis of Fatty Acids, Lipids and Membranes*, ed. O. Geiger (Cham: Springer International Publishing), 1–25.
- Sharma, P. K., Munir, R. I., Blunt, W., Dartailh, C., Cheng, J., Charles, T. C., et al. (2017). Synthesis and physical properties of polyhydroxyalkanoate polymers with different monomer compositions by recombinant *Pseudomonas putida* LS46 expressing a novel PHA SYNTHASE (PhaC116) enzyme. *Appl. Sci.* 7:242. doi: 10.3390/app7030242
- Sorokin, D. Y., Berben, T., Melton, E. D., Overmars, L., Vavourakis, C. D., and Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18, 791–809. doi: 10.1007/s00792-014-0670-9
- Sruamsiri, D., Thayanukul, P., and Suwannasilp, B. B. (2020). In situ identification of polyhydroxyalkanoate (PHA)-accumulating microorganisms in mixed microbial cultures under feast/famine conditions. *Sci. Rep.* 10, 3752–3752. doi: 10.1038/s41598-020-60727-7

- Surendran, A., Lakshmanan, M., Chee, J. Y., Sulaiman, A. M., Thuoc, D. V., and Sudesh, K. (2020). Can polyhydroxyalkanoates be produced efficiently from waste plant and animal oils? *Front. Bioeng. Biotechnol.* 8:169.
- Tan, I. K. P., Foong, C. P., Tan, H. T., Lim, H., Zain, N.-A. A., Tan, Y. C., et al. (2020). Polyhydroxyalkanoate (PHA) synthase genes and PHA-associated gene clusters in *Pseudomonas* spp. and *Janthinobacterium* spp. isolated from Antarctica. *J. Biotechnol.* 313, 18–28. doi: 10.1016/j.jbiotec.2020.03.006
- Tanikkul, P., Sullivan, G. L., Sarp, S., and Pisutpaisal, N. (2020). Biosynthesis of medium chain length polyhydroxyalkanoates (mcl-PHAs) from palm oil. *Case Stud. Chem. Environ. Eng.* 2:100045. doi: 10.1016/j.csee.2020.100045
- The UniProt C. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100
- Urschel, M. R., Kubo, M. D., Hoehler, T. M., Peters, J. W., and Boyd, E. S. (2015). Carbon source preference in chemosynthetic hot spring communities. *Appl. Environ. Microbiol.* 81, 3834–3847. doi: 10.1128/aem.00511-15
- Ver Eecke, H. C., Butterfield, D. A., Huber, J. A., Lilley, M. D., Olson, E. J., Roe, K. K., et al. (2012). Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. *Proc. Nat. Acad. Sci. U.S.A.* 109, 13674–13679. doi: 10.1073/pnas.1206632109
- Wang, B., Qin, W., Ren, Y., Zhou, X., Jung, M. Y., Han, P., et al. (2019a). Expansion of Thaumarchaeota habitat range is correlated with horizontal transfer of ATPase operons. *ISME J.* 13, 3067–3079. doi: 10.1038/s41396-019-0493-x
- Wang, L., Liu, Q., Wu, X., Huang, Y., Wise, M. J., Liu, Z., et al. (2019b). Bioinformatics analysis of metabolism pathways of archaeal energy reserves. *Sci. Rep.* 9:1034. doi: 10.1038/s41598-018-37768-0
- Wang, S., Chen, W., Xiang, H., Yang, J., Zhou, Z., and Zhu, M. (2016). Modification and potential application of short-chain-length polyhydroxyalkanoate (SCL-PHA). *Polymers* 8:273. doi: 10.3390/polym8080273
- Xiao, Z., Zhang, Y., Xi, L., Huo, F., Zhao, J. Y., and Li, J. (2015). Thermophilic production of polyhydroxyalkanoates by a novel *Aneurinibacillus* strain isolated from Gudao oilfield. *China. J. Basic Microbiol.* 55, 1125–1133. doi: 10.1002/jobm.201400843
- Zhong, H., Lehtovirta-Morley, L., Liu, J., Zheng, Y., Lin, H., Song, D., et al. (2020). Novel insights into the Thaumarchaeota in the deepest oceans: their metabolism and potential adaptation mechanisms. *Microbiome* 8:78. doi: 10.1186/s40168-020-00849-2
- Zhou, Z., Tran, P. Q., Kieft, K., and Anantharaman, K. (2020). Genome diversification in globally distributed novel marine *Proteobacteria* is linked to environmental adaptation. *ISME J.* 14, 2060–2077. doi: 10.1038/s41396-020-0669-4
- Ziemert, N., Alanjary, M., and Weber, T. (2016). The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* 33, 988–1005. doi: 10.1039/c6np00025h

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Vuong, Lim, Murphy, Wise, Whiteley and Kaur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.