Comparative metagenomics of PHA synthase genes in soil

by

Phuong Thi Ai Vahlis

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2017

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Polyhydroxyalkanoates (PHAs) are biopolymers produced naturally by bacteria. They are of considerable scientific interest as fundamental components of bacterial carbon metabolism and have biotechnological applications as potential bioplastics. To date, studies of PHA metabolism have focused on a restricted set of PHA-producing bacterial species. Therefore, the diversity of PHA-producing taxa and gene sequences, and the efficiency of existing primers to recognize PHA marker genes, is unclear. In this thesis, I report the first large-scale metagenomic analysis of PHA producing taxa through taxonomic and functional profiling of 45 soil metagenomes from a broad range of soil types (bulk and rhizosphere). From a total of 229,070 detected class I-III PHA synthase (*phaC*) genes, PHA-producing microbial communities were inferred and compared between soil environments, and the sequence diversity and primer efficiency for different classes of *phaC* genes was analyzed. Analysis revealed several main findings: 1) both known and novel PHA-producing taxa were inferred to contribute high proportions of *phaC* genes in environmental samples; 2) distinct shifts in the PHA-producer communities were observed both between soil types and between *phaC* classes; 3) *phaC*-containing species were detected at relatively higher abundance in rhizosphere soils implying a significant role for PHA storage in rhizobacteria; 4) existing primers did not adequately cover the sequence diversity of environmental homologs, and metagenomic diversity can be used to suggest modification that improve primer efficiency.

Acknowledgements

First and foremost, I wish to express my gratitude to my graduate supervisor Dr. Andrew Doxey for his guidance and patience throughout my Master of Science program. I started my research work with little background in Bioinformatics but Dr. Doxey has always generously given me all the support that I need to learn and grow in the field. I feel very fortunate to be part of Dr. Doxey's laboratory.

I also want to thank my committee members Dr. Brendan McConkey and to Dr. Trevor Charles for all the valuable insights on the topic of Polyhydroxyalkanoate producing bacteria. Additional thanks to all the lab members for their help and useful comments during my practice presentations.

I thank my parents and parents in law for their love and support.

Lastly, I wish to thank my husband Yevgeniy for being a loving, supportive husband and an inspiring scientist.

Dedication

To Yevgeniy and Noa

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

Polyhydroxyalkanoic acids (PHAs) are an extensively studied class of biodegradable, non-toxic, water-insoluble polyesters that accumulate intracellularly within a large number of microbial species ranging from bacteria to archaea to fungi (Jendrossek et al., 1996, Weiner 1997; Steinbuchel et al., 1998; Steinbuchel and Fuchtenbusch, 1998; Kim and Lenz, 2001; Hermann-Krauss et al, 2013; Cheng and Charles, 2016). PHA polymers accumulate as energy storage compounds by these microbial species in response to stress conditions where carbon is in excess relative to other nutrients such as nitrogen, sulfur and oxygen (Anderson and Dawes 1990). PHA's unique characteristics have been exploited for numerous industrial applications such as packings, medicine, pharmacy, and cosmetics (Anderson and Dawes, 1990; Braunegg et al. 1998; Zinn et al., 2001).

PHAs can be classified into different types based on the number and composition of repeating monomers. Polymers that contain three to five carbons such as polyhydroxybutyrate (PHB) or polyhydroxyvalerate (PHV) are short chain length (scl) PHAs. Polymers with six to fourteen carbon chains are considered medium chain length (mcl) PHAs. Short chain length PHAs are more commonly found in nature than mcl PHAs (Kim et al., 2007). Species of bacteria known to produce PHB include *Cupriavidus necator* (previously known as *Ralstonia eutropha*) and Rhizobia such as *Sinorhizobium, Rhizobium, Bradyrhizobium, Mesorhizobium, and Azorhizobium* (Trainer and Charles, 2006). The ability to accumulate mcl PHAs has been studied extensively in *Pseudomonas oleovorans, P. putida* and related species (Timm and Steinbuchel, 1990; Fuchtenbusch and Steinbuchel, 1999).

1

The PHB biosynthetic pathway is a 3-step-pathway starting from the precursor, Acetyl Coenzyme A (Acetyl-CoA) (**Figure 1**). Key enzymes in this pathway include β-ketothiolase, Acetoacetyl-CoA reductase, and PHA synthase/polymerase, which are products of the genes *phaA*, *phaB*, and *phaC*, respectively. The enzyme β-ketothiolase creates a carbon-carbon linkage for two Acetyl-CoA molecules through a Claisen condensation process to form AcetoAcetyl-CoA (Dawes, 1988; Steinbuchel and Schlegel, 1991). The enzyme Acetoacetyl-CoA reductase then reduces the compound formed in the previous step to D(-)-3-hydroxybutyryl-CoA, a monomeric precursor of PHA. The last step is the polymerization of PHB by PHB synthase, which catalyzes the release of Coenzyme A from the PHA monomeric precursor and forms an ester bonding between the new precursor and an existing polyester molecule (Steinbuchel and Schlegel, 1991) (**Figure 1**).

The microbial production of mcl PHAs has been investigated by numerous studies, which have resulted in the identification of two synthetic pathways (Huijberts et al., 1994; Langenbach et al., 1997; Qi et al., 1998 and 2000; Fiedler et al., 2000) (**Figure 1**). One route diverts the intermediates of the beta oxidation (Acyl-coa) to PHA synthesis when fatty acids are the main carbon sources. A second pathway, observed in some strains of *Pseudomonas* species such as *P. putida* and *P. aeruginosa* involves the incorporation of intermediates of *de novo* fatty acid synthesis (R-3-hydroxyacyl-ACP) to PHA when grown on non-related carbon sources such as gluconate (Hoffman et al., 2000a and b).
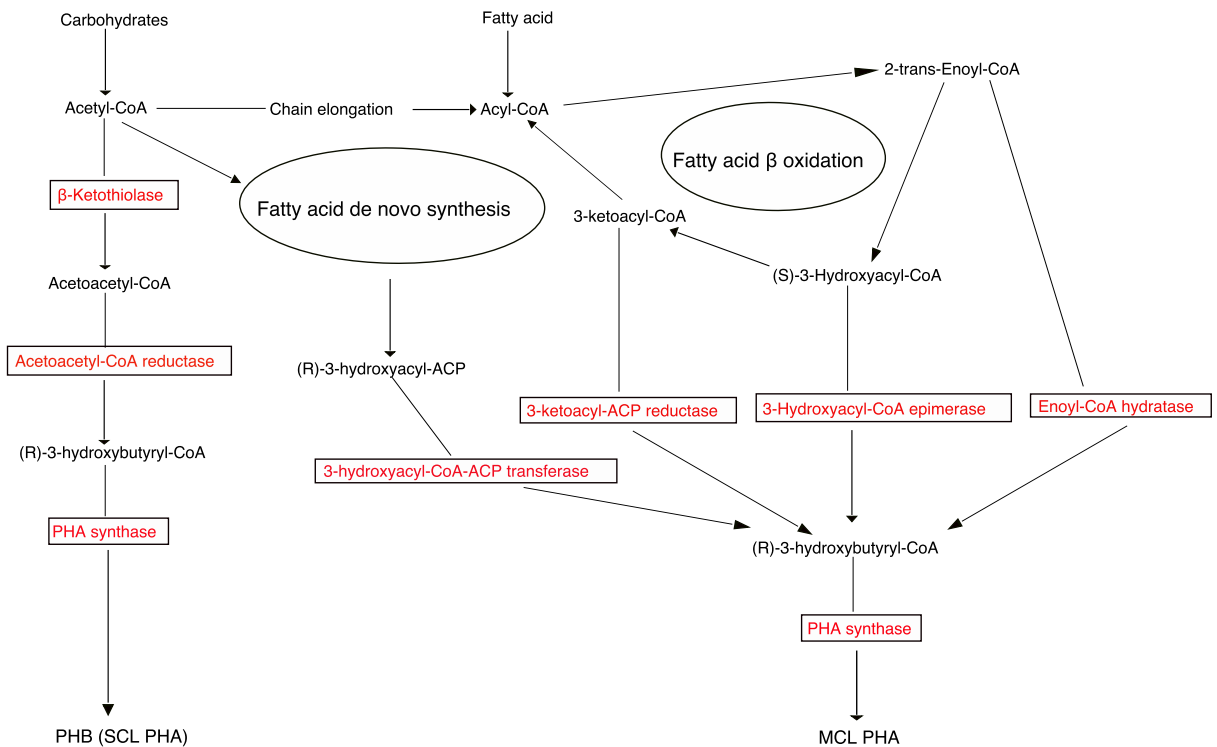
Figure 1. Biosynthesis pathways for the production of short chain length and medium chain length PHAs.

Currently, four different classes of PHA synthases exist, each forming a distinct protein family with a unique substrate specificity, domain and subunit composition. Class I PHA synthases form a single polypeptide chain possessing an N-terminal domain of unknown function as well as a C-terminal catalytic domain (Wittenborn et al., 2016), and possess activity against (R)-hydroxyacyl-CoA (3HAscl). As demonstrated by a recently solved crystal structure of a class I PHA synthase from *Cupriavidus necator* (Wittenborn et al., 2016), the catalytic domain of PHA synthase belongs to the α/β hydrolase fold super family (Rehm et al., 2003), similar to that seen in lipases. Class II synthases appear homologous to class I synthases but possess activity only against mcl (R)-3-hydroxyacyl-CoA (3HAmcl) and have been found in many members of

3

Pseudomonas. Class III PHA synthases are unique from both class I and II synthases by possessing a two-subunit structure (*phaC* and *phaE*), and have activity against both scl and mcl PHAs (Liebergesell and Steinbuchel, 1992; Liebergesell et al., 1994; Amara et al., 2002). *Allochromatium vinosum* which is formerly known as *Chromatium vinosum* is the representative species for class III PHA synthases (McCool and Cannon, 2001). Class IV PHA synthases also consist of two subunits (*phaC* and *phaR*) but have activity against scl PHAs only (Qi et al., 2000; Rehm 2003 and 2007). PHA synthase genes of class IV are commonly found in members of *Bacillus* genus and the representative species are *Bacillus megaterium* and *Bacillus cereus* (Tsuge et al., 2015).

Historical approaches for identifying PHA production activity have relied on phenotypic detection (e.g., staining with Sudan black B (Schlegel et al. 1970)), Nile blue A (Ostle and Holt 1982), or Nile red (Gorenflo et al. 1999; Spiekermann et al. 1999). Despite the sensitivity of these methods, they consume more time and require more effort to ensure that the testing environment is suitable for PHA to be built-up in bacterial cells. To overcome these problems, faster genotypic detection methods were developed. Two examples are polymerase chain reaction (PCR) technique used to develop degenerate primers for class I and class II synthase genes in studies of Sheu et al., 2000 and Solaiman et al., 2000, respectively. These primers were used to detect PHA synthase genes and screen for PHA producers in the environment. In their studies, primers from Sheu et al., 2000 were able to detect all 19 known PHA-positive bacteria and primers from Solaiman et al, 2000 demonstrated their ability to confirm *P. corrugata* to contain *phaC1/C2* genes and can accumulate mcl PHA.

In addition, numerous studies have attempted to identify novel genes involved in PHA metabolism, with the goal of exploring organisms beyond more typically studied, culturable bacteria. For example, Aneja et al. (2004) used phenotypic complementation to reveal novel PHB synthase genes (Aneja et al., 2004). In this study, phenotypic expression of mucoid colony morphology of *Sinorhizobium meliloti* strain Rm1021 on Yeast Mannitol Agar and heterologous complementation of a *Sinorhizobium meliloti phbC* mutant was used to confirm the presence of PHB synthase genes in *Bradyrhizobium japonicum.* This functional screening approach has been used to identify novel PHA metabolic genes in both soil and activated sludge (Wang et al., 2006). Metagenomic libraries constructed from these microbial communities were also used to complement a *Sinorhizobium meliloti bdhA* (3-hydroxybutyrate dehydrogenase) mutant, resulting in the identification of many clones that had regained PHA degradation activity. In a more recent study, Schallmey et al. (2011) used the same approach to identify and characterize novel PHA synthesis genes in a soil metagenomic library (Wang et al., 2006).

In addition to functional metagenomics approaches, bioinformatic data-mining of metagenomic datasets has also been an effective approach for identifying novel PHA synthases and circumventing the limitations of traditional isolation methods. Such approaches facilitate the study of both culturable and un-culturable bacteria from a large multitude of ecological niches (Sturgeon et al., 2013; Bibby and Peccia, 2013; Foong et al., 2014). One of the metagenomic approaches used in recent studies includes the use of degenerate primers in whole genome amplification to identify novel PHA synthase genes and subsequent genome walking to determine full-length coding sequences. The functionality of these novel genes has been

confirmed via the restoration of PHA accumulating ability in PHB-negative mutant strains of *Cupriavidus necator*. In 2014, Foong et al. (2014) applied this method to detect novel class I and II *phaC* genes from Japan Trech and Nankai Trough seawater. More recently, Tai et al. (2016) discovered novel class I and class II PHA synthases from metagenomes collected from Malaysian limestone hills and outcrops also by using the same technique.

The applications of PHAs as biodegradable thermoplastics with potential to replace conventional plastics provide strong motivations to search for novel PHA producing bacteria and biosynthetic enzymes. However, to date, no studies have made use of the large variety of existing soil metagenomic datasets to screen for novel PHA synthase genes. A large-scale bioinformatic survey of soil metagenomes is needed to characterize the taxa involved in PHA production in terrestrial environments, determine the biological differences between the PHA-producing community in different conditions, and also examine the effectiveness of existing PHA synthase primers.

In this study, we have conducted a bioinformatic survey and comparative analysis of novel class I, II and III PHA synthase genes in 45 available soil metagenomes. We identified and taxonomically classified PHA synthase genes, and compared their taxonomic profiles between bulk soils and rhizosphere soils to identify potential biologically meaningful differences in microbial PHA production. In addition, we compared the functional gene (PHA synthase) contribution from different taxa with their estimated community abundance based on a commonly used taxonomic marker gene (*rpoB*). Together this data revealed species that likely play a role in PHA production within the microbial community and their relative community

abundance. Finally, we examined the metagenomic data to measure the efficiency of existing degenerate primers for class I (Sheu et al., 2000) and class II PHA synthase genes (Solaiman et al., 2000). We hypothesized that the metagenomic data may help guide the design of improved primers that better represent that natural diversity of sequences present in soils and overcome biases of earlier primer designs based on PHA synthase genes from culturable bacteria, which potentially account for only a small fraction of environmentally prevalent bacteria. We then use metagenomics-guided primer design to optimize existing primers and improve their similarity to PHA synthase sequence diversity present in the environment.

Chapter 2- Materials and Methods

2.1 Bioinformatic prediction of PHA synthase classes

For the bioinformatic strategy to detect PHA synthases, hmmsearch and hmmscan from the HMMER package was used to search for matches to profile Hidden Markov Models (HMMs) representing class I-III PHA synthases. Profile HMMs were obtained from the TIGRFAM database (class IV was unavailable) as query models for metagenomic homology searches. TIGRFAM models were selected because, as demonstrated by a Conserved Domain Database (CDD) scan of representative class I-III sequences, they mapped effectively to the full-length multi-domain sequences of PHA synthases unlike single domain models (i.e., PFAM HMMs) (**Figure 2A**).
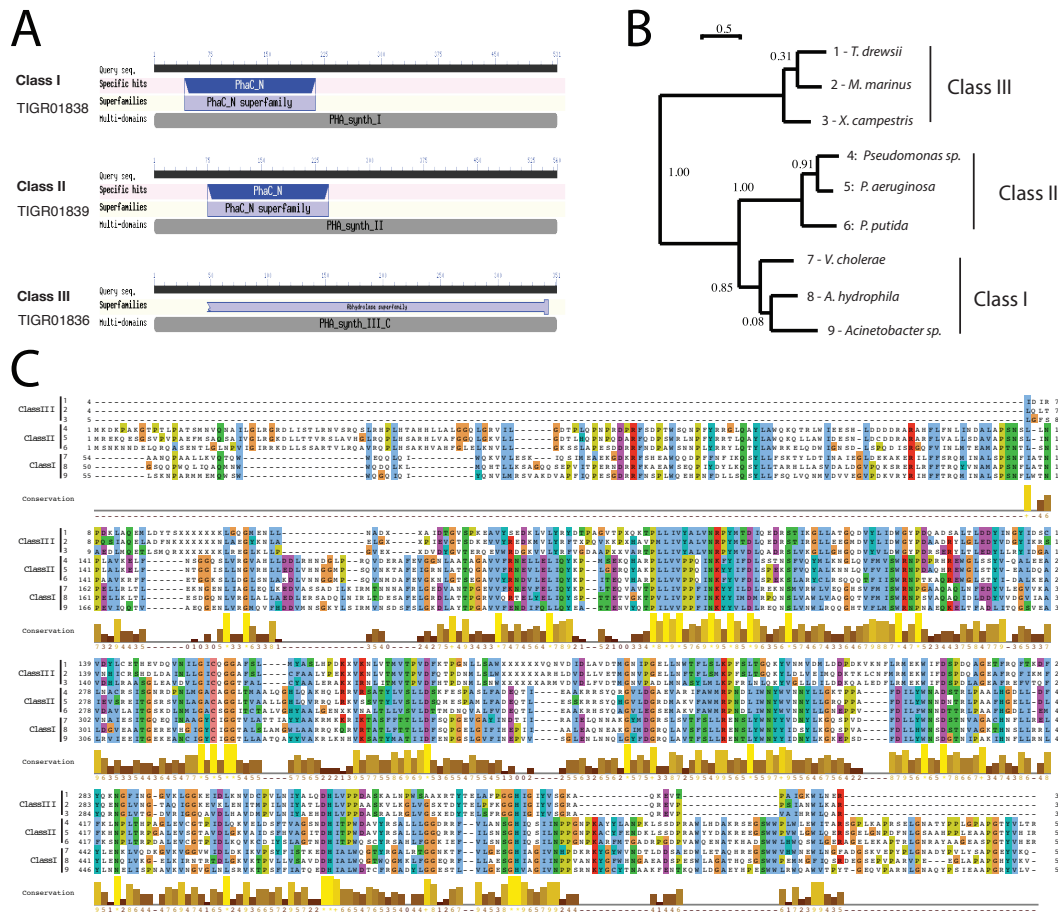
Figure 2. Bioinformatic analysis of representative class I, II, and III PHA synthases. Nine known PHA synthases from classes I-III were analyzed for their protein domain architecture (A) using the NCBI Conserved Domain Database. As shown above, TIGRFAM models mapped effectively to the representative full-length sequences. (B) A phylogenetic analysis of these representative sequences subdivided them into three subfamilies. As revealed in (C), classes I-III are homologous but show distinct patterns such an apparent deletion in the N and C-terminal domain of class III synthases relative to class I and II. NCBI accession numbers for the protein sequences labeled 1-9 are as follows: 1 - WP_007038931.1; 2 - WP_011712836.1; 3 - WP_011037305.1; 4 - WP_025340912.1; 5 - WP_070345320.1; 6 - WP_013974534.1; 7 - WP_053029769.1; 8 - AAL77053.1; 9 - WP_019384922.1.

9

Despite the phylogenetic relationship (**Figure 2B**) and clear pattern of sequence homology (**Figure 2C**) between class I-III sequences, the TIGRFAM models successfully differentiated sequences of the three classes of PHA synthases. To further investigate this, we examined the ability of TIGRFAM HMMs to accurately assign the correctly PHA synthase class to a test set of 18 known class I-III PHA synthases obtained from (Lu et al., 2008). TIGRFAM HMM identifiers for class I-III PHA synthases were TIGR01838, TIGR01839 and TIGR01836, respectively. To determine whether hmmsearch can accurately distinguish the overlapping domains in the HMM models for class I, II and III synthases, we ran a reciprocal search in which hmmscan was used to search for the identities of all the hits obtained from hmmsearch and confirmed how many of them were correctly identified. These 1-1 matches were used to assign PHA synthase class. In addition, we randomly sampled different fragment lengths (100, 50, 30, 20 and 10 amino acids) from this test set to better simulate the short reads obtained in metagenomic sequencing and their effect on accuracy. **Table 1** lists the fraction of correctly assigned "reads" for known class I-III PHA synthases. The results show that hmmsearch and hmmscan were no longer able to correctly detect hits with short simulated read lengths of 10 amino acids per sequence at an E-value threshold of 0.001. When the alignment length was increased to 20, 30, 50, 100 amino acids per sequence, the predictions were confirmed by hmmscan at 100% accuracy. In addition, we observed a dependence on the conservation of the region that the reads mapped to, in which in more conserved regions hmmsearch predictions were 100% accurate when the minimum length of the alignment is 20 amino acid per sequence.

In less conserved regions, the 100% accuracy of the predictions can only be obtained when the length of each sequence in the alignment is equal to or greater than 30 amino acids.

Table 1. Precision and sensitivity of HMM-based PHA synthase predictions using simulated read datasets of varying read lengths.

Read Length	Precision		Sensitivity	
	Conserved regions	Variable regions	Conserved regions	Variable regions
100	100%	100%	100%	100%
50	100%	100%	100%	100%
30	100%	100%	100%	100%
20	100%	No hits	100%	0%
10	No hits	No hits	0%	0%

2.2 Metagenomic datasets

Twenty-five bulk soil metagenomic datasets from MG-RAST and 20 rhizosphere metagenomic datasets from EBI metagenomics database were selected to search for novel PHA synthase genes. These bulk soil metagenomes were collected from different locations ranging from grasslands, temperate coniferous forest, tundra, desert and xeric shrubland and mangroves. The MG-RAST identifications of these soil metagenomic datasets are: 4477803.3, 4477872.3, 4477875.3, 4477877.3, 4477900.3, 4477902.3, 4477904.3, 4539064.3, 4541642.3, 4541645.3, 4541647.3, 4541649.3, 4541651.3, 4477805.3, 4477873.3, 4477876.3, 4477899.3, 4477901.3, 4477903.3, 4539063.3, 4541641.3, 4541644.3, 4541646.3, 4541648.3, 4541650.3. Wheat and cucumber rhizosphere metagenomes collected from the EBI metagenomics database are

SRR908275a and b, SRR908281a and b, SRR358565a and b, SRR908272a and b, SRR908276a and b, SRR908290a and b, SRR358608a and b, SRR908208a and b, SRR908273a and b, SRR908279a and b, SRR908291a and b.

2.3 Metagenomic taxonomic and functional profiling

The MetAnnotate bioinformatics pipeline (see **Figure 3**) was used to analyze all metagenomic datasets, and infer both their taxonomic and functional profiles. The Transeq program within the EMBOSS package (Rice et al., 2000) was used to first translate metagenomic reads into all six possible opening reading frames (ORFs). Subsequently, translated ORFs were searched using profile Hidden Markov Models (HMM search) from the HMMER package (Finn et al., 2011) to find homologs to query proteins of interest (i.e., PHA synthase protein families). The profile HMMs are more reliable than other methods such as BLAST (Altschul et al., 1997) because they are built from a collection of homologs and are therefore not biased by a single sequence query. They are also more sensitive than PSI-BLAST (Altschul et al., 1997) as they include insertion/deletion event probabilities, and not only position-specific scoring frequencies.

Profile HMMs for these enzymes were obtained from the TIGRfam database protein family (Haft et al., 2003). HMM searches were also performed against the NCBI RefSeq (Tatusova et al., 2013) database of non-redundant annotated reference sequences. The USEARCH program was then used to obtain the best RefSeq match to each metagenomic hit, and RefSeq hits were used to assign taxonomic information.
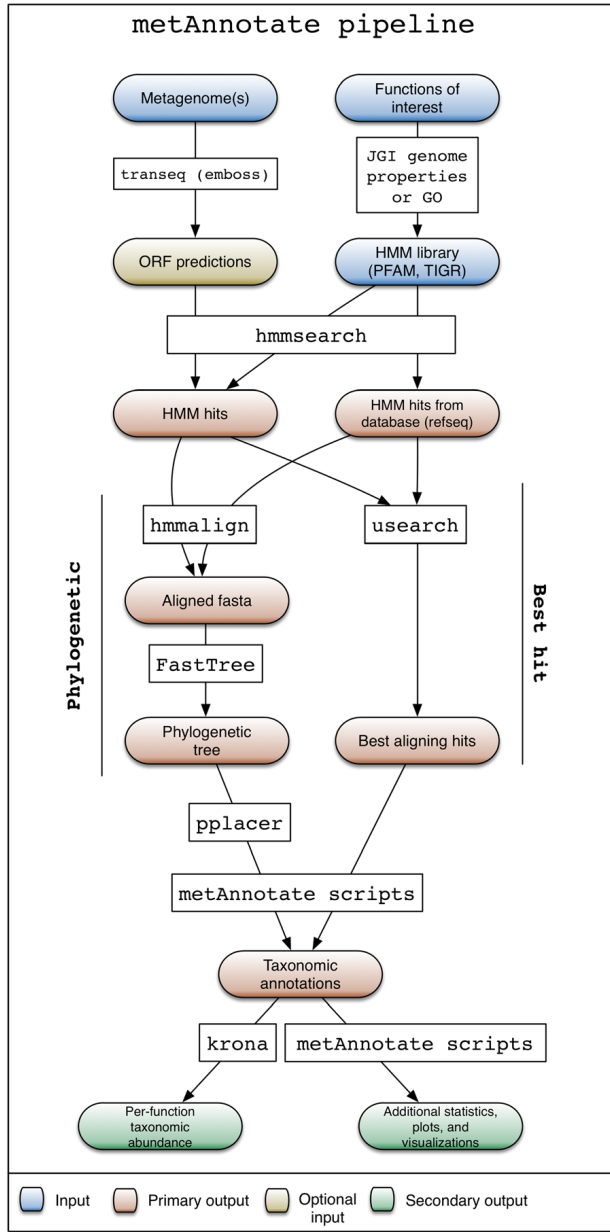
Figure 3. The MetAnnotate framework for taxonomic and functional profiling of metagenomes. Figure adapted from Petrenko et al. (2015).

2.4 Metagenomics-guided primer analysis and design

In order to identify and align primer regions for detected PHA synthases in soil metagenomes, a script (BackToDNA) was developed in the GO programming language. BackToDNA converts an amino acid alignment to its corresponding nucleotide alignment using the raw metagenomic reads. BackToDNA reads tabular output from BLASTp searches (protein queries) against the DNA reads (database) and selects hits with 100% identity and the maximum bit score. The script then processes the BLAST result to reverse translate the amino acid sequence (with gaps) into its corresponding DNA sequence. To accomplish this, the strand and frame must be accounted for, and gaps are replaced with three dash characters (see **Figure 4**).
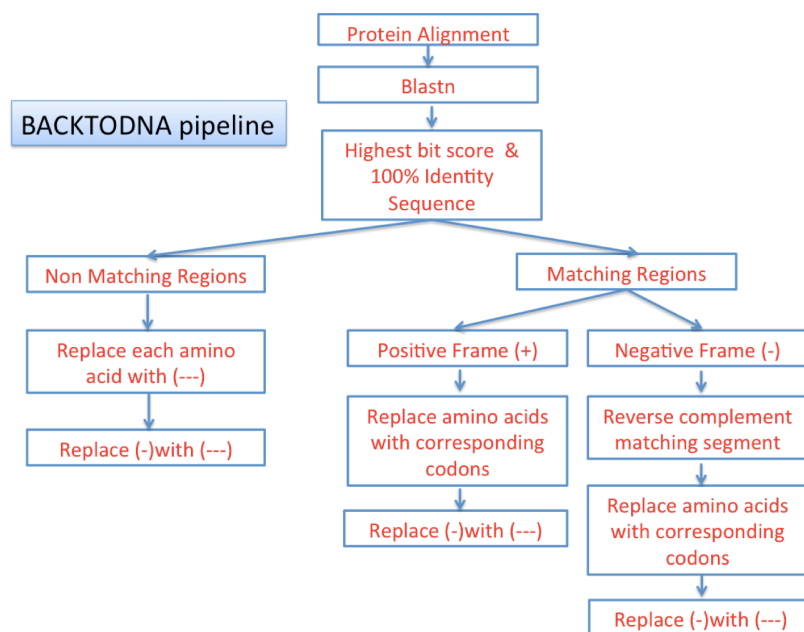
Figure 4. Flowchart of the BackToDNA method for generating metagenomics-guided primer alignments. The method begins with a protein alignment, and reverse translates it to a DNA-based alignment based on the original raw metagenomic reads.

Chapter 3 - Results

3.1 Taxonomic profiling of PHA synthases in soils

To detect novel class I, class II and class III PHA synthases, we retrieved 25 bulk soil metagenomic datasets from MG-RAST and 20 rhizosphere metagenomic datasets from the EBI metagenomics database. The bulk soil metagenomes span a broad range of environments including grasslands, desert and forest soils. Using the MetAnnotate pipeline (Petrenko et al., 2015), unassembled metagenomic reads were scanned for matches to class I-III PHA synthases and all hits were taxonomically annotated by sequence comparison to a reference protein database. Specifically, reads were translated into 6 possible open reading frames (ORFs) using EMBOSS Transeq (Rice et al., 2000), ORF datasets were scanned for matches to TIGRFAM profile HMMs of class I-III PHA synthases (see Methods) using hmmsearch from the HMMER package (Finn et al., 2011), and the hits were taxonomically classified based on the top RefSeq matches computed using usearch (Edgar, 2010). Metagenomic hits that could not be matched to any Refseq sequences were labeled as "Unclassified".

The metagenomic annotation pipeline identified 128,006 metagenomic reads associated with class I PHA synthase genes, 72,364 reads associated with class II PHA synthase genes and 28,700 reads associated with class III PHA synthase genes. However, there was some prediction overlap between different PHA synthase classes (see **Figure 5**). Therefore, to improve prediction accuracy, all PHA synthase hits were scanned against the full TIGRFAM HMM library (Haft et al., 2003), which either verified the initial class prediction or disambiguated the class in cases where a read matched multiple classes (**Figure 5**). After reciprocal match filtering,

110,872 (86.6%) of class I genes, 32,575 (45.0%) of class II genes, and 23,165 (80.7%) of class III genes were verified. Consistent with a phylogenetic tree of representative PHA synthases (**Figure 2B**), the largest extent of overlapping hits occurred between class I and class II genes (**Figure 5**). Only these reads verified through reciprocal matches were selected for further analyses.

MetAnnotate
hmmsearch

PhaC synthase   →   Metagenomic hits
                 ←

Hmmscan
TIGRFAM
library
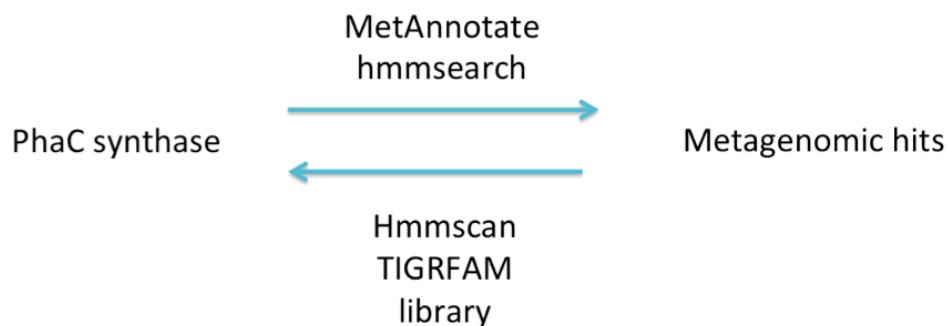
Hmmsearch		Class I	Class II	Class III
		128,006	72,364	28,700
Hmmscan	Class I	**110,872**	39,476	5,257
	Class II	14,813	**32,575**	278
	Class III	2,321	313	**23,165**

Figure 5. Number of hits and confusion matrix for predicted class I-III PHA synthases in soil metagenomes. As shown above, TIGRFAM HMMs were used to search metagenomic reads for hits to class I-III synthases and hits were in turn scanned with the full TIGRFAM HMM library in order to detect "top reciprocal matches".

Next, the taxonomic profiles of the metagenomic PHA class I-III synthases were inferred using MetAnnotate (Petrenko et al., 2015). **Figures 6-8** illustrate the most abundant taxa contributing class I, II and III PHA synthase genes in bulk soil and rhizosphere metagenomes (also see **Tables 2-4**). Below, we explore these taxa and compare predictions to existing literature on known class I-III PHA synthase producers.

3.1.1 Class I PHA synthases

In rhizosphere metagenomes, detected class I PHA synthase reads affiliated most with the *Sphingomonadaceae* (18.3%), *Rhizobiaceae* (11.6%), *Caulobacteraceae* (8.3%), and *Comamonadaceae* (7.1%) (**Figure 6, Table 2**). In bulk soils, however, *Bradyrhizobiaceae* (13.2%), *Rhodospirillaceae* (7.7%), *Burkholderiaceae* (7.4%), and *Rhodocyclaceae* (6.2%) were the most dominant phyla contributing class I PHA synthases (**Figure 6**, **Table 2**).

These results are largely consistent with existing literature on known PHA producers. For example, the accumulation of PHB as an energy storage polymer is well established in members of the *Rhizobiaceae* and *Bradyrhizobiaceae* families (Tombolini and Nuti, 1989; Tombolini et al., 1995; Quelas et al., 2013). In addition, PHB biosynthetic activity has been demonstrated in *Caulobacteraceae* (Poindexter 1981; Qi and Rehm, 2001), in *Rhodobacter* and *Rhodospirilum* (Hustede et al., 1992; Liebergesell et al., 1991), and in the *Comamonadaceae* (Sudesh et al., 1998; Yee et al., 2012).

17
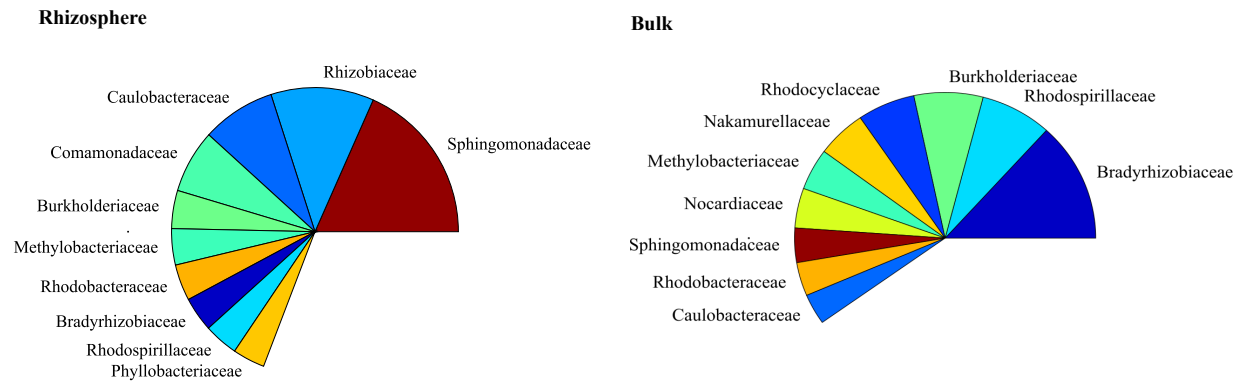
Figure 6. Taxonomic distribution of class I PHA synthases detected in 45 rhizosphere and bulk soil metagenomes.

At the same time, the abundance of *phaC* class I genes affiliated with the *Sphingomonadaceae*, which dominated the contribution of *phaC* genes in rhizosphere soils, is surprising since they have been relatively understudied in the context of PHA production. One study by Godoy et al. (2003) has shown that members of the *Sphingomonadaceae* (*Sphingopyxis* species) are capable of accumulating scl PHA. More recently Cheng and Charles (2016) identified by functional metagenomics a *Sphingomonas*-related sequence that complemented a *phaC* negative strain of *Pseudomonas putida.* The considerable abundance of *Sphingomonadaceae*-related *phaC* class I genes in metagenomes suggests an underappreciated role for these taxa in the production of PHA compounds in the rhizosphere environment.

Table 2. Per-taxa abundance of class I *phaC* genes and *rpoB* genes detected in 45 rhizosphere and bulk soil metagenomes.

Rhizosphere	Fraction *phaC*	Raw Data *phaC*	Fraction *rpoB*	Raw Data *rpoB*	Class
Closest Homolog Family					
Sphingomonadaceae	0.183475	5878	0.084985	38029	*Alphaproteobacteria*
Rhizobiaceae	0.116178	3722	0.05188	23215	*Alphaproteobacteria*
Bradyrhizobiaceae	0.038799	1243	0.03635	16266	*Alphaproteobacteria*
Rhodobacteraceae	0.040672	1303	0.025872	11577	*Alphaproteobacteria*
Caulobacteraceae	0.08306	2661	0.024359	10900	*Alphaproteobacteria*
Comamonadaceae	0.071105	2278	0.020139	9012	*Betaproteobacteria*
Phyllobacteriaceae	0.036333	1164	0.014533	6503	*Alphaproteobacteria*
Rhodospirillaceae	0.03858	1236	0.010535	4714	*Alphaproteobacteria*
Burkholderiaceae	0.042763	1370	0.010461	4681	*Betaproteobacteria*
Methylobacteriaceae	0.040703	1304	0.005596	2504	*Alphaproteobacteria*
Bulk Soil	Fraction *phaC*	Raw Data *phaC*	Fraction *rpoB*	Raw Data *rpoB*	Class
Closest Homolog Family					
Bradyrhizobiaceae	0.132553	8817	0.051804	32518	*Alphaproteobacteria*
Sphingomonadaceae	0.038246	2544	0.016468	10337	*Alphaproteobacteria*
Rhodospirillaceae	0.076777	5107	0.016302	10233	*Alphaproteobacteria*
Rhodobacteraceae	0.037224	2476	0.015236	9564	*Alphaproteobacteria*
Caulobacteraceae	0.033931	2257	0.009071	5694	*Alphaproteobacteria*
Burkholderiaceae	0.074101	4929	0.00683	4287	*Betaproteobacteria*
Rhodocyclaceae	0.061849	4114	0.006627	4160	*Betaproteobacteria*
Nocardiaceae	0.044079	2932	0.005268	3307	*Actinobacteria*
Methylobacteriaceae	0.046018	3061	0.005146	3230	*Alphaproteobacteria*
Nakamurellaceae	0.053415	3553	0.00116	728	*Actinobacteria*

3.1.2 Class II PHA synthases

Among the Class II PHA synthases, members of the *Pseudomonadaceae, Geodermatophilaceae,* and *Nocardiaceae* dominated the metagenomics hits detected, and these top three taxa were consistent between rhizosphere and bulk soils, although they differed in relative abundance.

Pseudomonadaceae contributed the majority of genes in rhizosphere soils (55%) and roughly 1/6 (16.1%) of the genes in bulk soils (**Figure 7**). Indeed, members of the *Pseudomonadaceae* are well known mcl PHA producers (Timm and Steinbuchel, 1992). *Pseudomonas* spp. are known to possess two medium chain length PHA synthase genes, *phaC1* and *phaC2*, (Huisman et al., 1989 and Timm et al., 1994) and these genes in conjunction with the flanking *phaZ* gene encoding PHA depolymerase, have been shown to confer the ability to synthesize mcl PHA (Timm and Steinbuchel, 1992).
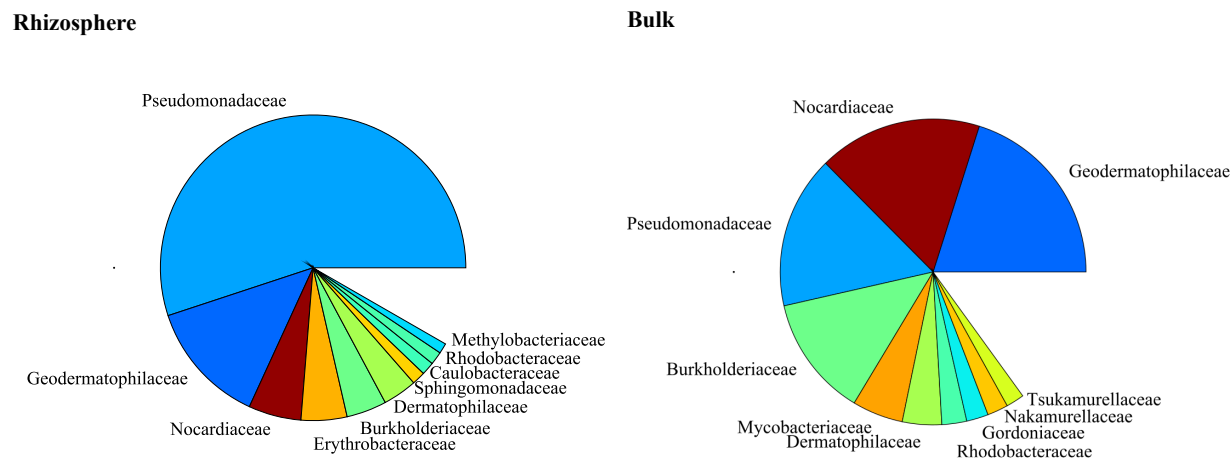
Figure 7. Taxonomic distribution of class II PHA synthases detected in 45 rhizosphere and bulk soil metagenomes.

Members of the *Nocardiaceae* family were also detected as major contributors to class II *phaC* genes in both rhizosphere (5.6%) and bulk soil (17.3%) (**Figure 7**). Indeed, previous studies have demonstrated mcl PHA producing activity by the *phaC* gene (phaCNc) of *Nocardia corallinia* (Timm and Steinbuchel, 1992; Hall et al., 1998).

Interestingly, several phyla contributing class II PHA synthases are relatively understudied in the context of PHA production. The families *Geodermatophilaceae* (13.1% in rhizosphere, 20.1% in bulk soil) and *Dermatophilaceae* (3.6% in rhizosphere, 4.2% in bulk soil) contributed significant proportions of class II PHA genes (**Figure 7**, **Table 3**), and yet the ability of these taxa to accumulate mcl PHAs is to our knowledge less well known. In one study, Liu et al. (2002) discovered a novel species (*Kineosphaera limosa)* within the *Dermatophilaceae* that was demonstrated to accumulate PHA. Our analysis identified *phaC* class I reads affiliated with *Kineosphaera* as well as three additional genera (*Modestobacter, Blastococcus, Geodermatophilus*), suggesting that mcl PHA production may be more environmentally abundant and widespread within the *Dermatophilaceae* family than currently thought. Metagenomic hits for these taxa had identities to reference sequences ranging from 55%-100% percentage similarity. Similarly, although of lower abundance, the families *Nakamurellaceae*, *Tsukamurellaceae*, and *Gordoniaceae* were all predicted to contribute class II *phaC* gene homologs in soil metagenomes, but there is little if any literature knowledge on mcl PHA production in these species.

Table 3. Per-taxa abundance of class II *phaC* genes and *rpoB* genes detected in 45 rhizosphere and bulk soil metagenomes.

Rhizosphere	Fraction *phaC*	Raw Data *phaC*	Fraction *rpoB*	Raw Data *rpoB*	Class
Closest Homolog Family					
Sphingomonadaceae	0.014112	86	0.084985	38029	*Alphaproteobacteria*
Rhodobacteraceae	0.013784	84	0.025872	11577	*Alphaproteobacteria*
Caulobacteraceae	0.013784	84	0.024359	10900	*Alphaproteobacteria*
Pseudomonadaceae	0.550706	3356	0.023733	10620	*Gammaproteobacteria*
Burkholderiaceae	0.042501	259	0.010461	4681	*Betaproteobacteria*
Nocardiaceae	0.055793	340	0.007594	3398	*Actinobacteria*
Methylobacteriaceae	0.01083	66	0.005596	2504	*Alphaproteobacteria*
Geodermatophilaceae	0.130784	797	0.005377	2406	*Actinobacteria*
Erythrobacteraceae	0.048572	296	0.003097	1386	*Alphaproteobacteria*
Dermatophilaceae	0.036101	220	0.00147	658	*Actinobacteria*
Bulk	**Fraction *phaC***	**Raw Data *phaC***	**Fraction *rpoB***	**Raw Data *rpoB***	**Class**
Closest Homolog Family					
Rhodobacteraceae	0.026233	653	0.015236	9564	*Alphaproteobacteria*
Mycobacteriaceae	0.053833	1340	0.008099	5084	*Actinobacteria*
Geodermatophilaceae	0.201149	5007	0.007002	4395	*Actinobacteria*
Burkholderiaceae	0.128154	3190	0.00683	4287	*Betaproteobacteria*
Nocardiaceae	0.173108	4309	0.005268	3307	*Actinobacteria*
Gordoniaceae	0.0231	575	0.004618	2899	*Actinobacteria*
Pseudomonadaceae	0.161337	4016	0.001995	1252	*Gammaproteobacteria*
Dermatophilaceae	0.04158	1035	0.001818	1141	*Actinobacteria*
Nakamurellaceae	0.022055	549	0.00116	728	*Actinobacteria*
Tsukamurellaceae	0.019163	477	0.000421	264	*Actinobacteria*

Lastly, *Burkholderiaceae* is unique in its apparent contribution of both class I and class II synthases in soil metagenomes (**Figure 6**, **Figure 7**). This is consistent with previous literature, which has demonstrated *Burkholderiaceae* to possess both types (Cheng and Charles, 2016).

3.1.3 Class III PHA synthases

Metagenomic hits to class III PHA synthases show more dramatic patterns in terms of extreme taxonomic abundance and variation between environments (**Figure 8, Table 4**). *Xanthomonadaceae* dominated the contribution of class III genes in rhizosphere soils, making up 66.3% of class III reads. In bulk soils, *Bradyrhizobiaceae* (18.6%), and *Methylobacteriaceae* (11.6% in bulk soil) contributed the most class III genes.

The extreme abundance of *Xanthomonadaceae*-associated class III PHA synthases is surprising given the almost complete lack of literature on PHA production in this family. Many members of *Xanthomonadaceae* are well known as plant pathogenic and industrial producers of the exopolysaccharide, xanthan gum (Poli et al., 2011). A *phaC* gene from *Xanthomonas campestris* has been annotated computationally and included in a few previous studies (Aneja et al., 2009 and Amara and Moawad (2011). In addition, as shown in **Figure 9**, a gene cluster containing *phaC* and other PHB synthesis genes can be identified within the genomes of *Xanthomonas* species. However, experimental work to demonstrate PHA production and characterize this putative PHA synthase and the production pathway in this taxon is incomplete.
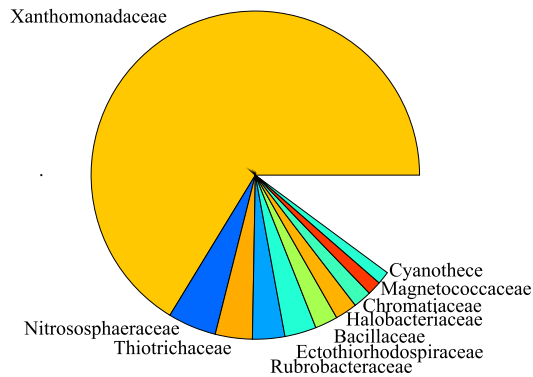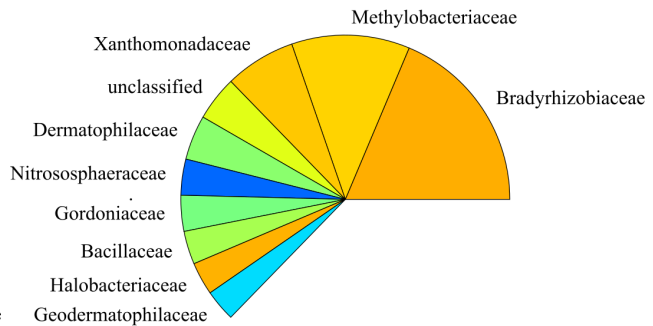
Figure 8. Taxonomic distribution of class III PHA synthases detected in 45 rhizosphere and bulk soil metagenomes.
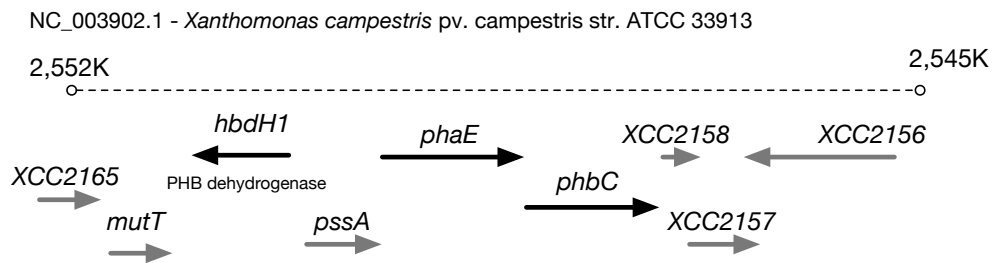
Figure 9. A gene cluster from *Xanthomonas campestris* pv. campestris str. ATCC 33913 containing putative *phaC*, *phaE* and *bdhA* genes.

The soil-associated archaeal family *Nitrososphaeraceae* also contributed a moderate amount of class III genes in both bulk (3.5%) and rhizosphere (4.8%) samples (**Figure 8, Table 4**). Most members of *Nitrososphaeraceae* from phylum *Thaumarchaeota* phylum possess genes that encode PHA synthases (Spang et al., 2012; Poli et al., 2011). In Zhalnina et al. (2014) the

24

presence of class III PHA synthase encoding genes (*phaC* and *phaE*) in *Candidatus Nitrososphaera evergladensis* was also confirmed.

Table 4. Per-taxa abundance of class III *phaC* genes and *rpoB* genes detected in 45 rhizosphere and bulk soil metagenomes.

Rhizosphere	Fraction *phaC*	Raw Data *phaC*	Fraction *rpoB*	Raw Data *rpoB*	Class
Closest Homolog Family					
Xanthomonadaceae	0.662931	5383	0.052604	23539	*Gammaproteobacteria*
Rubrobacteraceae	0.031773	258	0.007381	3303	*Actinobacteria*
Chromatiaceae	0.01798	146	0.004541	2032	*Gammaproteobacteria*
Bacillaceae	0.022414	182	0.003526	1578	*Bacilli*
Ectothiorhodospiraceae	0.031158	253	0.003066	1372	*Gammaproteobacteria*
Thiotrichaceae	0.036453	296	0.001124	503	*Gammaproteobacteria*
Nitrososphaeraceae	0.047783	388	0.000447	200	*Nitrososphaeria*
Magnetococcaceae	0.013793	112	0.000364	163	*Alphaproteobacteria*
Cyanothece	0.012685	103	0.000051	23	*Cyanobacteria*
Halobacteriaceae	0.021675	176	0.000011	5	*Halobacteria*
Bulk	Fraction *phaC*	Raw Data *phaC*	Fraction *rpoB*	Raw Data *rpoB*	Class
Closest Homolog Family					
Bradyrhizobiaceae	0.186574	2807	0.051804	32518	*Alphaproteobacteria*
Xanthomonadaceae	0.069724	1049	0.009203	5777	*Gammaproteobacteria*
Geodermatophilaceae	0.03104	467	0.007002	4395	*Actinobacteria*
Methylobacteriaceae	0.116251	1749	0.005146	3230	*Alphaproteobacteria*
Gordoniaceae	0.035294	531	0.004618	2899	*Actinobacteria*
Bacillaceae	0.032768	493	0.001819	1142	*Bacilli*
Dermatophilaceae	0.044001	662	0.001818	1141	*Actinobacteria*
unclassified	0.044001	662	0.001276	801	NA

Nitrososphaeraceae	0.035427	533	0.00073	458	*Nitrososphaeria*
Halobacteriaceae	0.032303	486	0.000003	2	*Halobacteria*

3.2 Sequence novelty of metagenomic *phaC* genes

To explore the novelty of *phaC* sequences identified in the soil metagenome survey, we examined their protein sequence identities to their top database (NCBI RefSeq) matches. In general, class I metagenomic hits possessed the highest similarity to reference proteins (**Figure 10**), with some clear skew towards culturable organisms that have been widely studied and sampled such as the *Rhizobiaceae* and *Bradyrhizobiaceae*. In rhizosphere soils, the average percentage identities to *Rhizobiaceae* ranged from 80-90% and in bulk soil, and the percentage identity of tops hits to *Bradyrhizobiaceae* ranged from 70-90%.

Analysis of sequence novelty for class II and III hits however, revealed more interesting patterns (**Figure 11, Figure 12**). That is, large variation in sequence novelty was observed both in taxa associated with PHA production as well as those that have not been as highly studied in this functional context. For example, in bulk soil, detected class II synthases affiliated with the *Pseudomonadaceae* were quite novel in sequence (~50-65% identities to the database), even though members of the *Pseudomonadaceae* are well known medium chain length PHA producers (**Figure 11**). On the other hand, the detected genes from taxa that are less commonly associated with PHA production such as *Dermatophilaceae* and *Geodermatophilaceae* were more similar to database sequences. Similarly, in rhizosphere soils, class III hits affiliated with the *Xanthomonadaceae*, which is not a widely studied taxa in terms of its PHA synthase

26

potential, were the least novel (the most similar to database sequences, ~75-85% identity) (**Figure 12**).

Ultimately, the sequence novelty analysis reveals that existing RefSeq genomes encode uncharacterized and potentially unnoticed PHA synthases in organisms that may not yet be associated with PHA synthesis. The identification of these organisms in soil metagenomes as dominant contributors of PHA synthase genes in soil environments places them into a functional context and provides motivation for future experimental work investigating the role of these species in PHA synthesis.
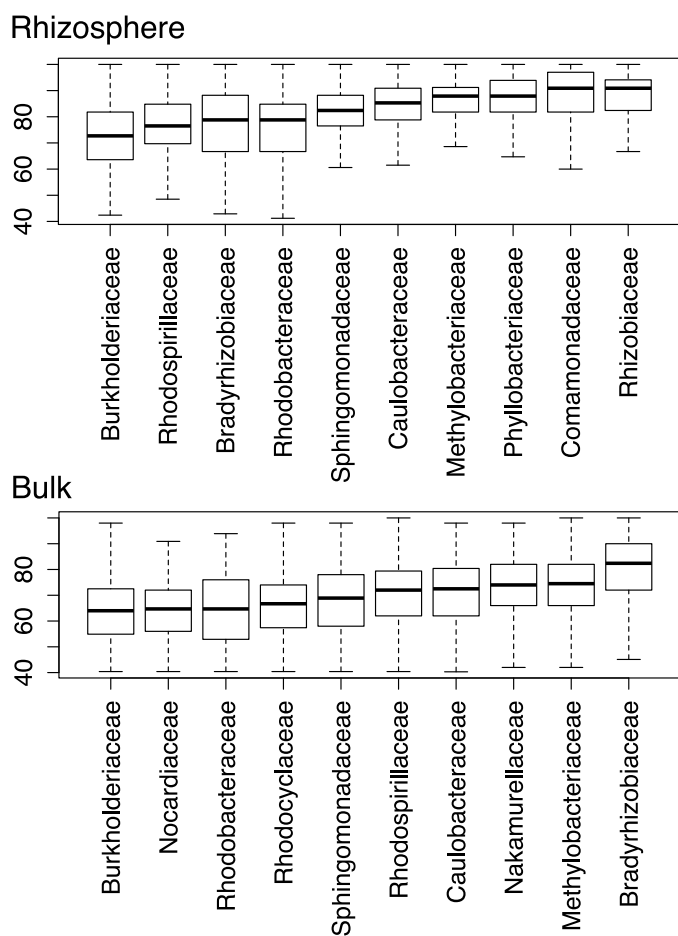
Figure 10. Sequence novelty of detected class I PHA synthases in 45 soil metagenome. The boxplots above show the distributions of sequence identities of detected class I PHA synthase genes for each taxa relative to their closest homologs in the reference database (NCBI RefSeq).
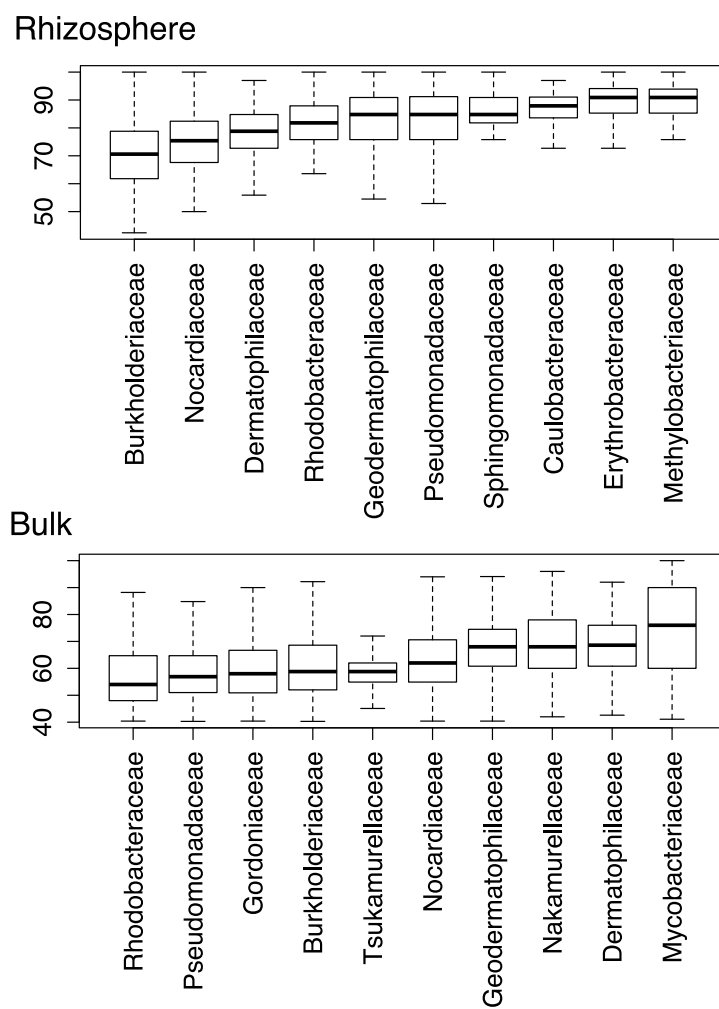
Figure 11. Sequence novelty of detected class II PHA synthases in 45 soil metagenome. The boxplots above show the distributions of sequence identities of detected class II PHA synthase genes for each taxa relative to their closest homologs in the reference database (NCBI RefSeq).
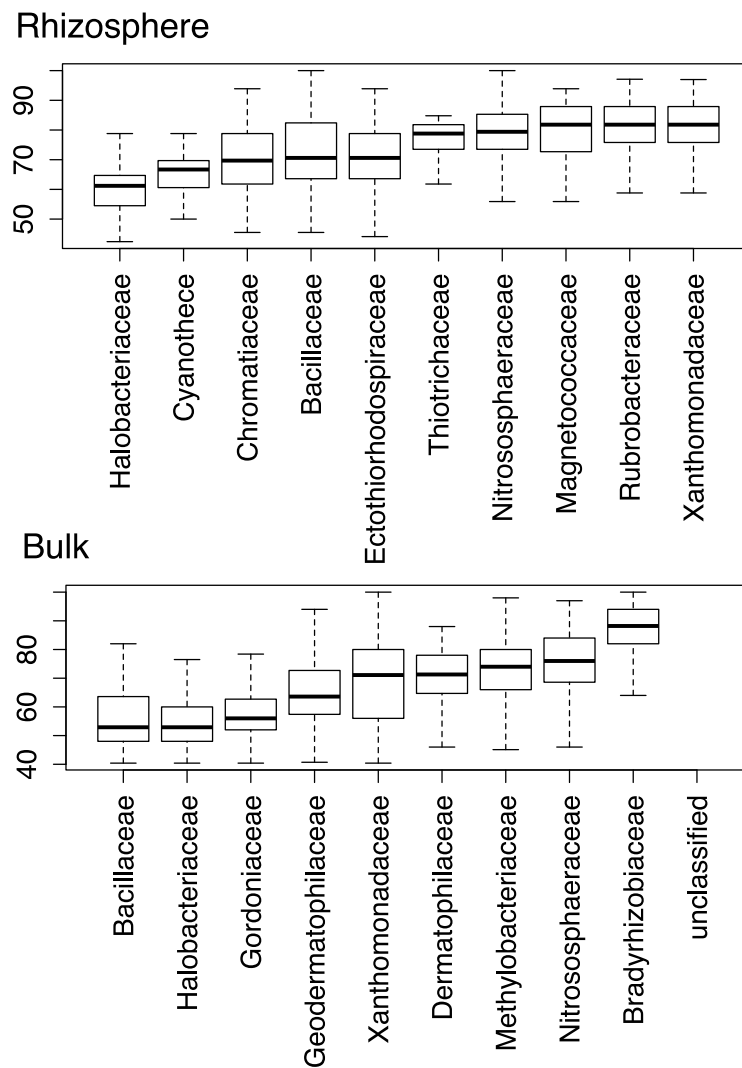
Figure 12. Sequence novelty of detected class III PHA synthases in 45 soil metagenome. The boxplots above show the distributions of sequence identities of detected class III PHA synthase genes for each taxa relative to their closest homologs in the reference database (NCBI RefSeq).

Interestingly, the novelty level of *phaC* metagenomic sequences from the bulk soil metagenomes is significantly higher than that in the rhizosphere as shown in **Figure 13**, which is consistent with the idea that soil metagenomes contain enormous diversity and community richness (Howe et al., 2014).
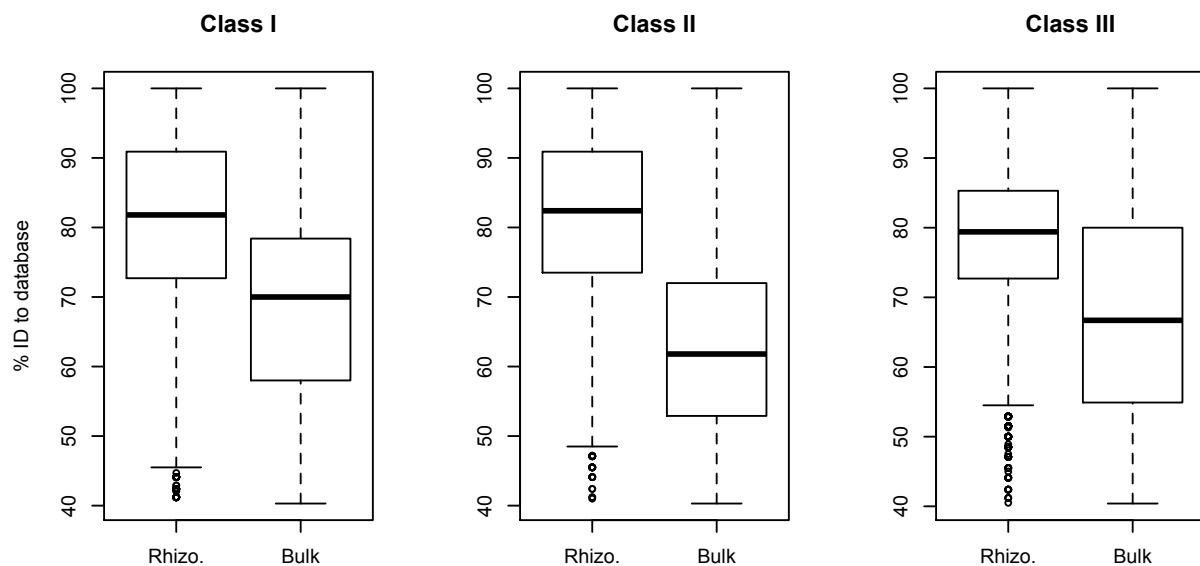
Figure 13. Distributions of sequence novelty of detected class I, II and III PHA synthases in rhizosphere versus bulk soil.

3.3 Comparison of PHA producer taxonomic profiles between bulk soil and rhizosphere metagenomes

Next, we compared *phaC* taxonomic profiles between bulk soil and rhizosphere metagenomes to identify biologically relevant differences. Using the pheatmap package within R, heat maps were generated to visualize the relative abundance of PHA synthases affiliated with different

taxonomic groups (family level) for classes I-III and for both rhizosphere and bulk soils (**Figures 14, 15, 16**).

3.3.1 Class I PHA synthases

Comparative analysis revealed a considerably higher abundance of class I genes affiliated with *Sphingomonadaceae*, *Rhizobiaceae*, *Caulobacteraceae* and *Comamonadaceae* families in rhizosphere soil than bulk soil (**Figure 14**). On the other hand, class I synthase gene contributions by *Bradyrhizobiaceae*, *Methylobacteraceae*, *Burkhoderiaceae*, *Rhodospirilaceae* were significantly higher in bulk soil (**Figure 14**). Taxonomic profiles of class I PHA synthases also subdivided further into distinct patterns according to soil region/type. For example, grassland and forest soils showed subtle differences between each other and desert soils showed the most dramatically different pattern (**Figure 14**). On the other hand, very little difference was observed between the *phaC* taxonomic profiles from cucumber and wheat-associated rhizosphere soils.
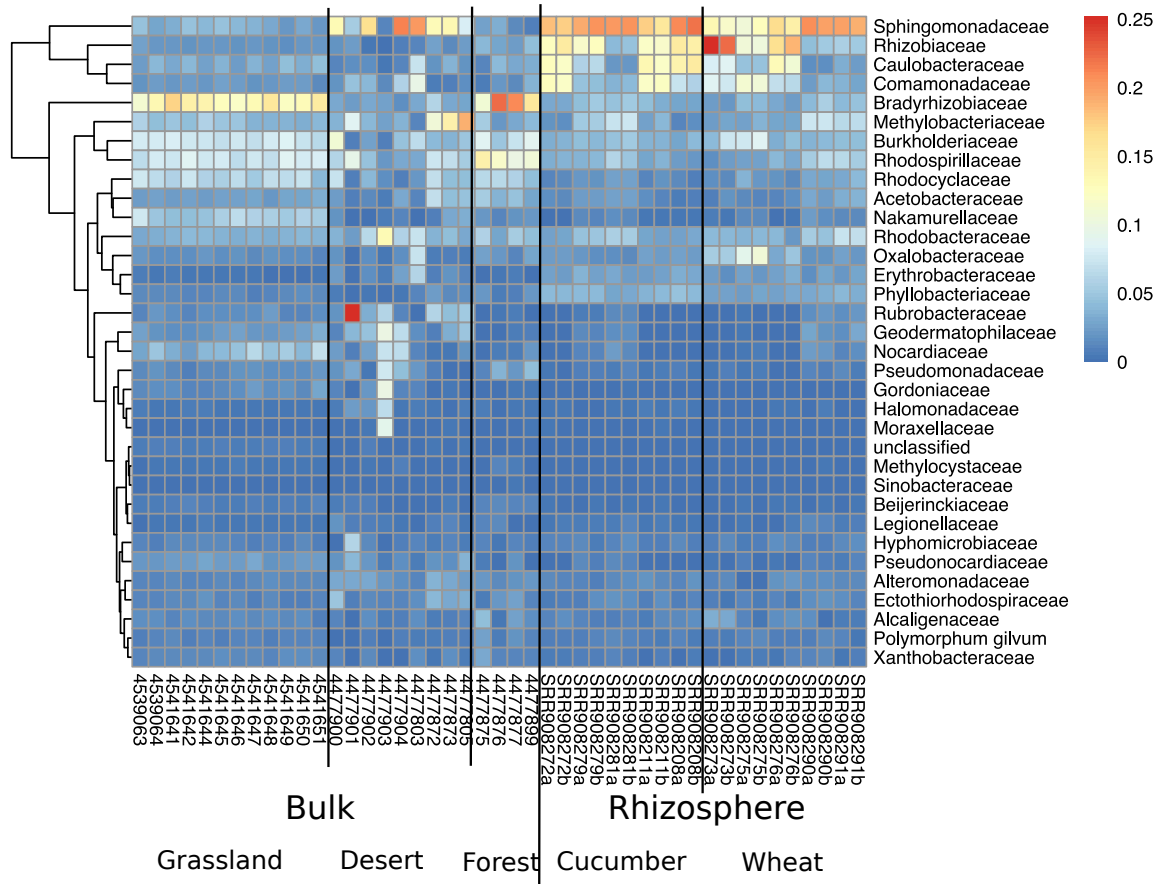
Figure 14. Taxonomic abundance heat map of detected class I PHA synthases assigned to each taxa in 45 metagenomes from rhizosphere versus bulk soils. Heatmaps were generated using the pheatmap() package in R and the y-axis (taxonomy) was clustered by Euclidean distances using the hclust() function.

The higher enrichment of PHA synthase class I gene in rhizosphere by *Rhizobiaceae, Comamonas* and *Sphingomonas* can potentially be explained by their roles as plant growth promoting bacteria in soil (Liu et al., 2006; Glick 2012, Ahemad and Kibret, 2014). The close associations of these genera with plant hosts may also explain their relative increase in the

rhizosphere. Their relative increase may be tied to the role of the host plants in supplying these bacteria with a carbon source for PHB accumulation.

Many members of *Bradyrhizobium* have also been mentioned in existing literature as plant growth promoting bacteria (Ahemad and Kibret, 2014). It was hypothesized that a higher PHA synthase gene enrichment by this genus will be found in rhizosphere soil than bulk soil. However, the comparative taxonomic profile of class I synthase genes demonstrates otherwise (**Figure 14**). There are a few possible explanations for this result. First of all, *Bradyrhizobium* species are slow growing bacteria in terms of carbohydrate consumption (Jordan 1982). Although fast growing rhizobia such as *Sinorhizobium meliloti* or *R. leguminosarum* can metabolize a wide range of carbohydrates, slow growing bacteria such as *Bradyrhizobium japonicium* and *B. lupini* can not use disaccharides, trisaccharides and organic acids for growth (Stowers 1985). Jordan (1982) suggested that slow-growing rhizobia seem to have the ability to consume large quantities of aromatic growth substrates. Therefore, slow growth rhizobia may be less dependent on plant roots in the rhizosphere soil for a carbon supply. In addition, some *Bradyrhizobium* strains are non-nodulating rhizobia, whose ability to survive and flourish in bulk soil has been documented previously (Gano-Cohen et al., 2016). In the same study, non-nodulating *Bradyrhizobium* strains demonstrated an ability to compete with nodulating strains for nodule interior and perhaps carbon supply. This might imply that they can use their *phaC* genes to accumulate PHAs in the rhizosphere soil region and continue to thrive after migrating to the bulk soil.

The abundance heat map (**Figure 14**) reveals a higher relative abundance of class I synthase genes affiliated with *Burkholderia* in bulk soil than in rhizosphere soil. A possible explanation for this result is that many members of the beta-proteobacteria have free living lifestyles, are tolerant to pH fluctuations or high salt extremes, and have loose relationship with plants (Gyaneshwa et al., 2011). It is possible that they can accumulate PHB in the rhizoshere soil region and continue to grow in bulk soil environments where there are less threat from other nodulating competitors and predators such as protozoa.

The accumulation and degradation of PHB by *Rhodospirillum rubum* from the *Rhodospirillaceae* family has been studied extensively. As a nitrogen fixer and producer of cytokinins, *R. rubum* is a well-established plant growth promoting bacteria (Glick 2012). The abundance heat map illustrates that members of this family contribute more class I *phaC* genes in bulk soil than in rhizosphere (**Figure 14**). There are several possible explanations for this result. *R. rubum* is a highly versatile organism and can grow both aerobically and anaerobically and has many ecological preferences (Cohen-Bazire and Kunisawa, 1963; Madigan et al., 2004 ; Reslewic et. al, 2005;). In addition, *R. rubum* is capable of photosynthesis under low oxygen conditions thanks to the presence of bacteriochlorophylls and carotenoids in the cell membrane. The fact that bacteriochlorophylls can absorb higher wavelength (800-925 nm) than plant's chlorophyll a (660-680 nm) means that *R. rubum* can use more energy from the light spectrum to accumulate energy compound by itself. This strain of bacteria has been shown to be energy-independent from plants and therefore, can thrive in bulk soil where there are less competitions and threats from other species present in rhizosphere soil.

3.3.2 Class II PHA synthases

The heat map for class II *synthase* genes (**Figure 15**) illustrates a much lower taxonomic diversity than that seen for class I genes (**Figure 14**), with only a few species dominating class II gene contribution in bulk soils and even fewer in rhizosphere. In general, the detected abundance of class II *phaC* genes from the *Pseudomonadaceae* family is much higher in rhizosphere soils, while the abundance of class II genes from *Geodermatophilaceae*, *Burkhoderiaceae* and *Norcardiaceae* are much higher in bulk soil.

Members of the *Pseudomonadaceae* are well known for their ability to accumulate medium chain length PHA, a product of the class II *phaC* genes. Many species of *Pseudomonadaceae* are also widely established as plant growth promoting bacteria. Pseudomonads play roles in helping plants obtain iron under iron-limiting conditions (Raaijmakers et al., 1994; Sharma and Johri, 2003; Masalha et al., 2000; Schalk et al., 2001), benefit plants during drought (Sandhya et al., 2010), and other conditions. The important roles of *Pseudomonas* species in promoting plant growth provides a potential explanation for the relative abundance of *Pseudomonas* class II synthase genes in rhizosphere soils.
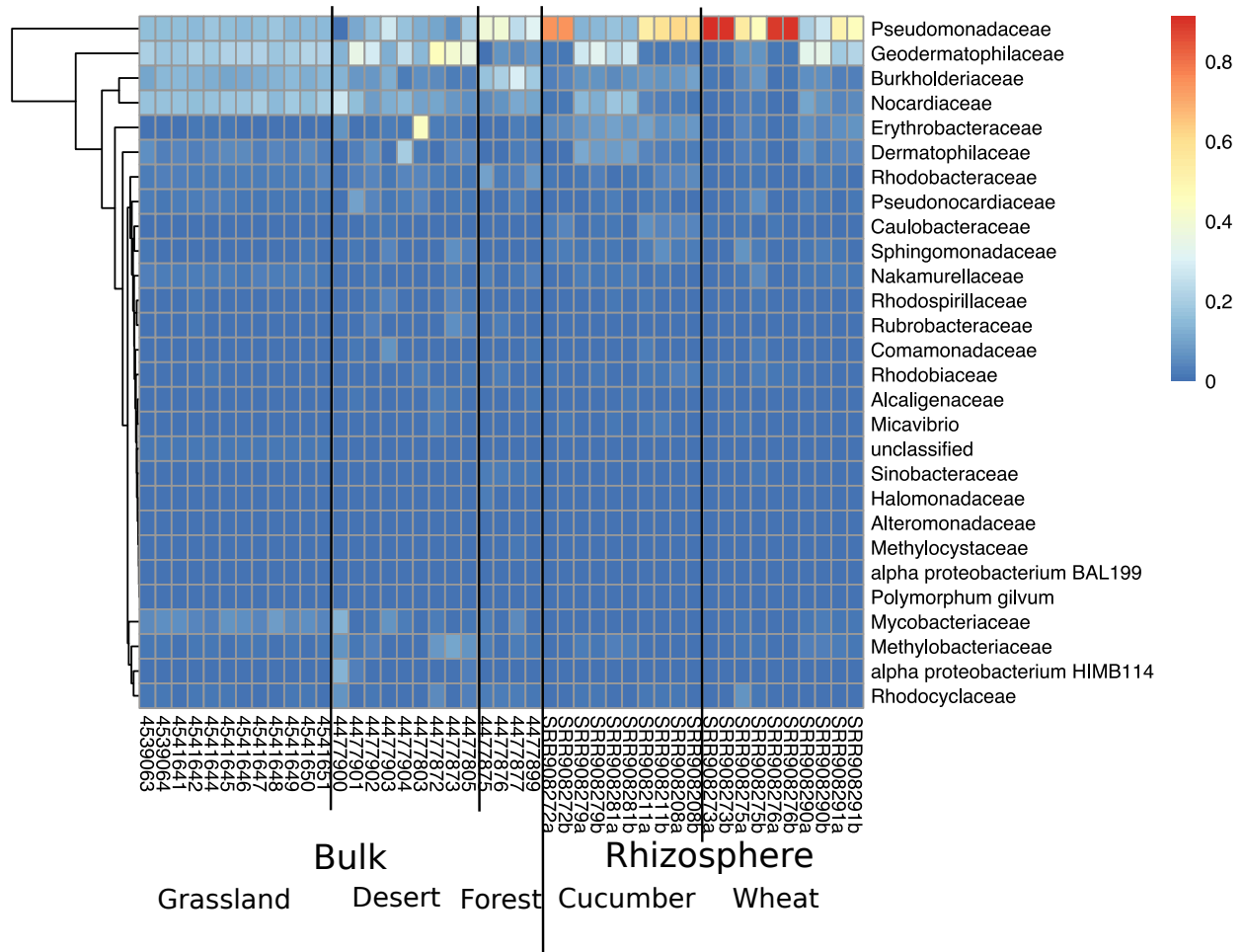
Figure 15. Taxonomic abundance heat map of detected class II PHA synthases assigned to each taxa in 45 metagenomes from rhizosphere versus bulk soils. Heatmaps were generated using the pheatmap() package in R and the y-axis (taxonomy) was clustered by Euclidean distances using the hclust() function.

As mentioned earlier, the ability of accumulating PHAs by members of genus *Geodermatophilus* has not been reported. Our study implies that these bacteria contribute a significant proportion of class II synthase genes in natural soil environments. In general, a greater abundance was observed in grassland and desert bulk soil (**Figure 15**), but a subset of individual rhizosphere

metagenomes, both from cucumber- and wheat-associated soils, soils also showed elevated abundance. Previous studies suggest that members of this genus can tolerate harsh environments such as arid soil, rock vanish in desert, UV light, and heavy metals, although some can be found in rhizosphere soil and lake sediments as well (Hezbri et al., 2015; Ivanova et al, 2010, Gtari et al., 2012; Rainey et al., 2005; Montero-Calasanz et al., 2014). This environmental versatility explains why this species was found to contribute class II genes in high abundance across a range of environmental soil types. A similar explanation has been provided for genera *Burkholderia* with regards to PHA synthase gene earlier in this section.

3.3.3 Class III PHA synthases

The taxonomic distribution of metagenomic class III PHA synthases (**Figure 16**) is perhaps even more restricted than that for class II genes. Class III PHA synthases from the *Xanthamonadaceae* family dominated all rhizosphere soils (both cucumber and wheat-associated), while *Bradyrhizobiaceae* and a few additional taxa dominated bulk soils, which also showed more variable patterns (**Figure 16**).

The *Xanthomonas* genus is known for its numerous species that cause plant diseases such as the phytopathogen, *X. campestris*. Their strong association with plant hosts provides an explanation for their relative abundance in rhizosphere soils. However, interestingly, in addition to dominating the rhizosphere soils, class III PHA synthases affiliated with *Xanthomonas* were also detected at high abundance in a subset of desert and forest metagenomes (**Figure 16**), suggesting a more widespread ecological niche.

Members of the *Rubrobacteraceae* family belongs to phylum *Actinobacteria*. They can be thermophilic, radio-tolerant and halo-tolerant (Suzuki et al., 1988; Carreto et al., 1996; Chen et al., 2004) . Consistent with an ability to survive in extreme environments, class III PHA synthases affiliated with the *Rubrobacteraceae* were detected at high abundance particularly in desert soils (**Figure 16**). At the species level, *Rubrobacter xylanophilus* was detected as the closest homolog for these metagenomic hits. However, there is no existing literature describing the presence of PHA synthase genes in these bacteria.
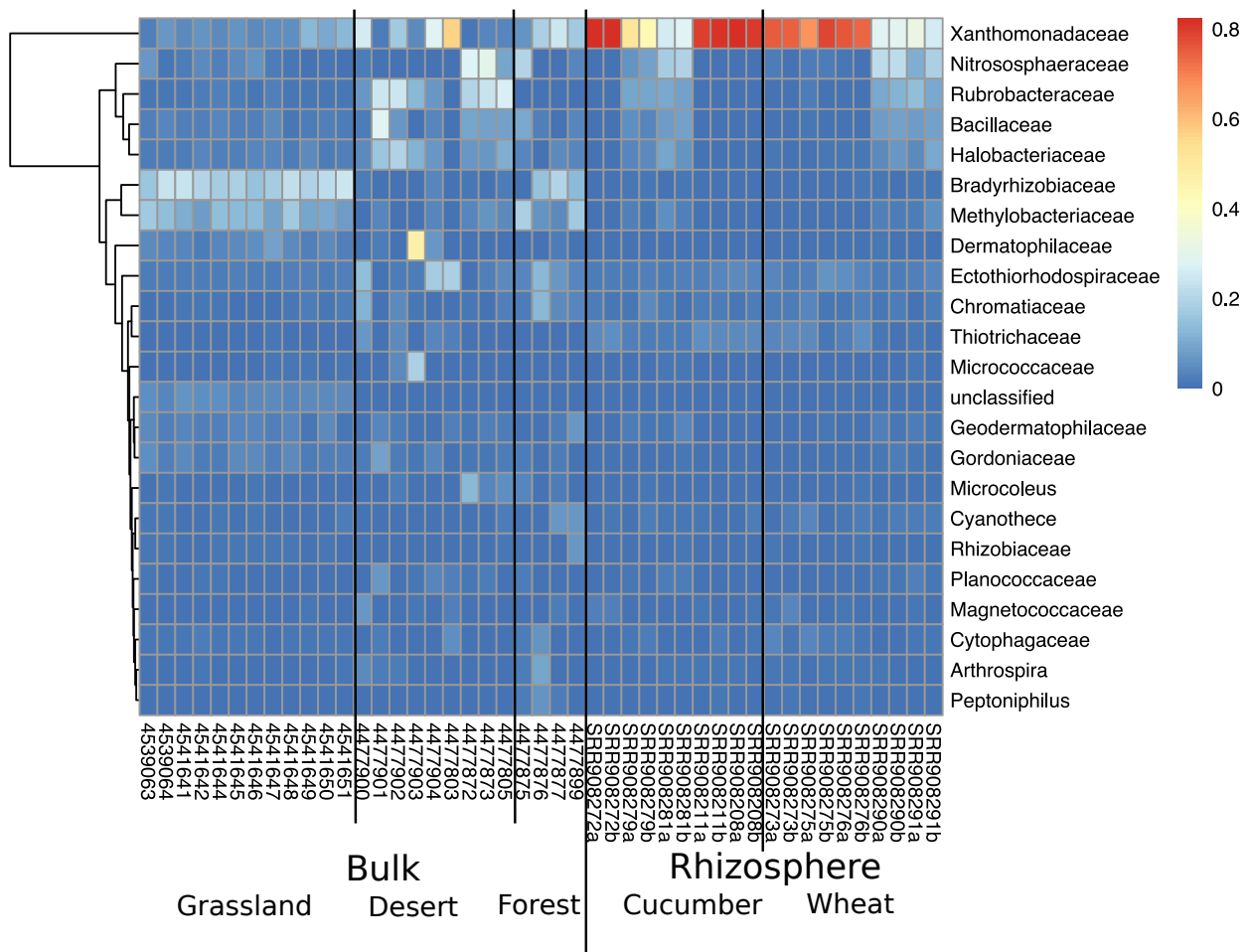
Figure 16. Taxonomic abundance heat map of detected class III PHA synthases assigned to each taxa in 45 metagenomes from rhizosphere versus bulk soils. Heatmaps were generated using the pheatmap() package in R and the y-axis (taxonomy) was clustered by Euclidean distances using the hclust() function.

Allochromatium vinosum is a well-studied anoxygenic purple sulfur bacteria from the phylum, gammaproteobacteria, and family, *Chromatiaceae*. It is both a sulfide and thiosulfate oxidizer. Because this family is primarily associated with aquatic environments (freshwater, brackish water or marine), it is quite interesting that our MetAnnotate analyses detected *Allochromatium vinosum* as the closest homolog for many of our soil metagenomes. As members of purple sulfur bacteria group, bacteria of *Chromatiaceae* are capable of doing photosynthesis, which can mean they are less dependent on plant for carbon supply. Perhaps, this can also imply their ability to survive well in bulk soil and contribute a higher enrichment of PHA synthase gene here than in rhizosphere soil.

3.4 Analysis of the community abundance of top PHA synthase contributors

In previous sections, the taxonomic distributions of PHA synthases (classes I-III) have been explored. However, how do these functional gene-based taxonomies correspond with the community abundance of the same organisms? To investigate this question, the per-taxon *phaC* gene abundance was compared with *rpoB* (housekeeping taxonomic marker) gene abundance.
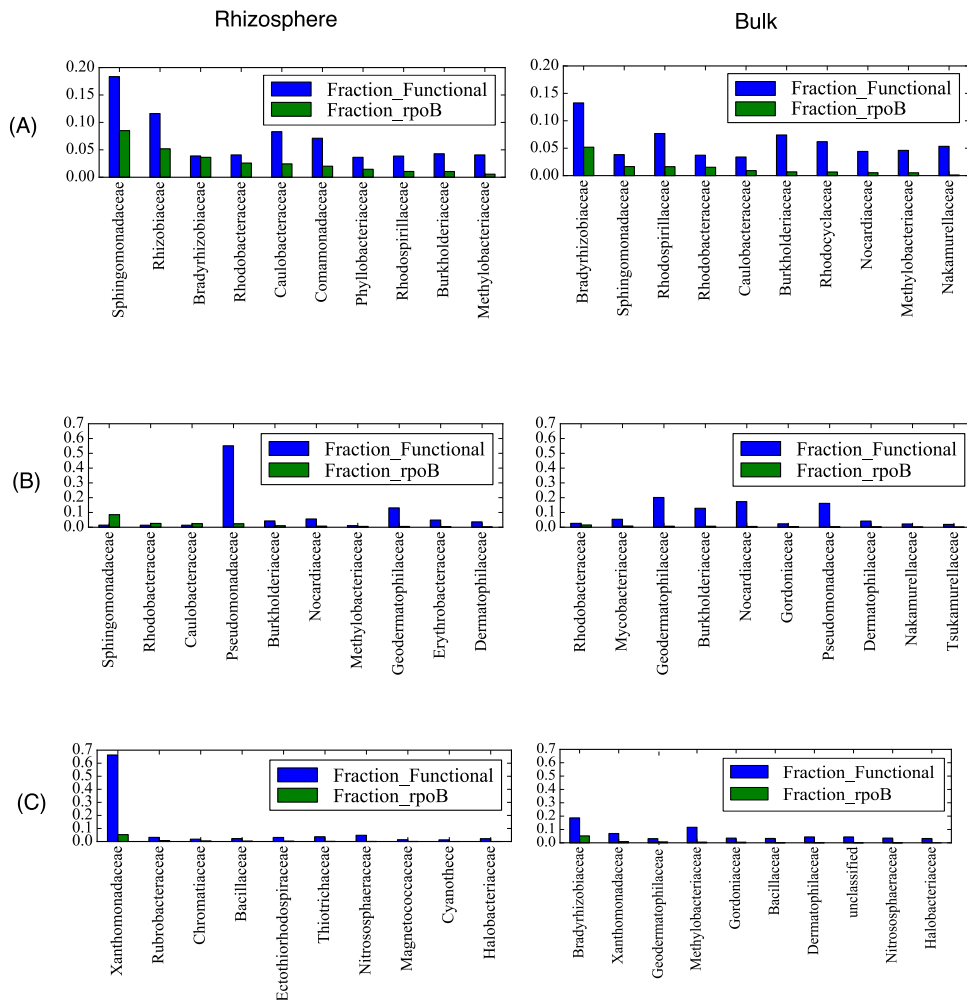
Figure 17. Proportion of *rpoB* genes (relative community abundance) and proportion of class I (A), class II (B), and class III (C) PHA synthase genes for the top PHA synthase contributing taxa in soil and rhizosphere metagenomes.

Figure 17 illustrates the top ten most abundant bacteria based on *rpoB* proportional abundance versus their PHA synthase gene abundance. Different patterns are revealed for different classes of synthase genes. For instance, in rhizosphere soils, abundant community members (taxa with

41

high *rpoB* abundance) such as *Sphingomonadaceae* and *Rhizobiaceae* also show high PHA synthase class I abundance. In bulk soil, however, this pattern is less pronounced. Class I contributors in bulk soils tend to be at relatively lower community abundance, as indicated by the high level of class I genes contributed by rare taxa such as *Methylobacteriaceae* and *Nakamurellaceae*. In addition, the class I producer community is similar between bulk and rhizosphere, sharing many of the same taxonomic groups. Together, this suggests that although class I PHA producing species exist in bulk soil, the same organisms become enriched in rhizosphere environments, and that the rhizosphere may therefore select more strongly for scl producers than bulk soils. This pattern is quantified in **Figure 18**, and is statistically significant according to a Wilcoxon Rank Sum test ($P = 0.043$). Since this analysis was restricted to the top ten families associated with PHA production may be biased. Therefore the analysis was repeated by examining the community abundance (estimated as the total proportion of *rpoB* genes) for all genera contributing *phaC* genes (**Figure 19**). In these plots, the community abundance of PHA-producing taxa as consistently higher in the rhizosphere than in bulk soil. This pattern is consistent for all three classes of PHA synthases. The detection of significant differences in PHA producing communities between bulk and rhizosphere is consistent with a large body of literature on the role of scl and mcl PHAs in plant-associated microbes (see Thesis Discussion).

In contrast to class I genes, the community abundance versus class II gene abundance in rhizosphere showed some unique patterns. For instance, the major class II contributors are not necessarily the most abundant organisms. *Pseudomonadaceae* is the main contributor of class II synthase genes. Although it is the fourth most abundant organism in the

community, its functional gene proportion vastly outweighs its proportional abundance in the community. In bulk soil, the community of class II synthase contributors is more diverse and spread out among several taxonomic groups including the *Geodermatophilaceae*, *Burkhoderiaceae*, *Norcadiaceae*, *Pseudomonadaceae* and other taxa with lower *rpoB* fractions. Ultimately, the weaker correspondence between class II abundance and community abundance suggests that there may be a relatively weaker role for class II genes in determining the growth or abundance of an organism both in soils and rhizosphere communities than that of class I genes.

Class III synthase gene analyses show similar patterns to class II analyses with differences in PHA contributing taxa. In rhizosphere soil, *Xanthomonadaceae* is the main class III synthase contributor and also has an extremely high PHA synthase fraction of almost 70%. However, again the class III gene contribution for *Xanthomonadaceae* vastly outweighs its community abundance (5.3%). Following from this, one may speculate whether scl PHA production by class III *phaC* genes gives *Xanthomonas* a competitive advantage in the rhizosphere community. In bulk soil, the class III gene contributing microbial community shows more diversity with a wider range of taxa including *Bradyrhizobiaceae*, *Xanthomonadaceae*, *Methylobacteriaceae*, *Gordoniaceae*, *Dermatophilaceae* and other taxa with lower community abundance.
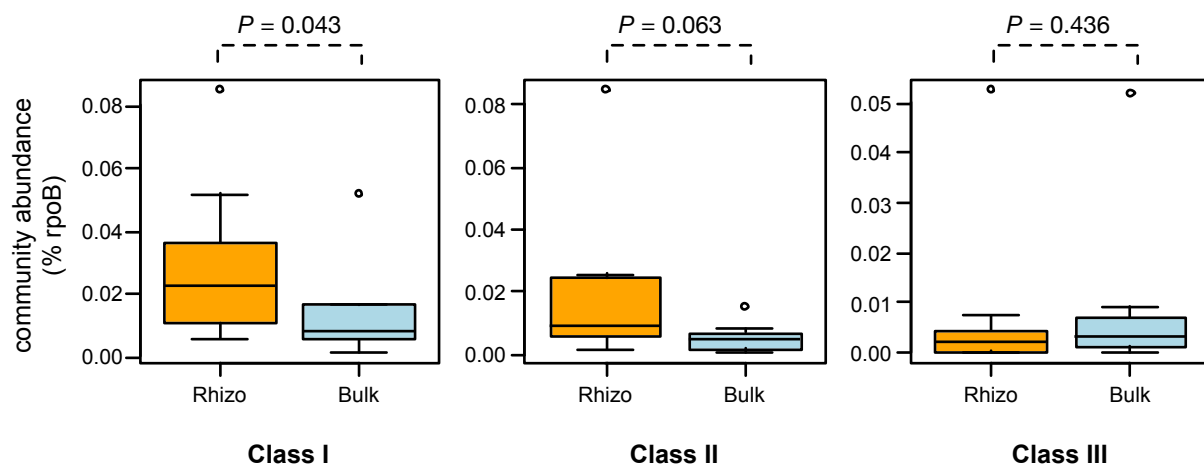
Figure 18. Community abundance of top 10 PHA-producing taxa associated with class I, class II, and class III *phaC* in bulk soil versus rhizosphere metagenomes.
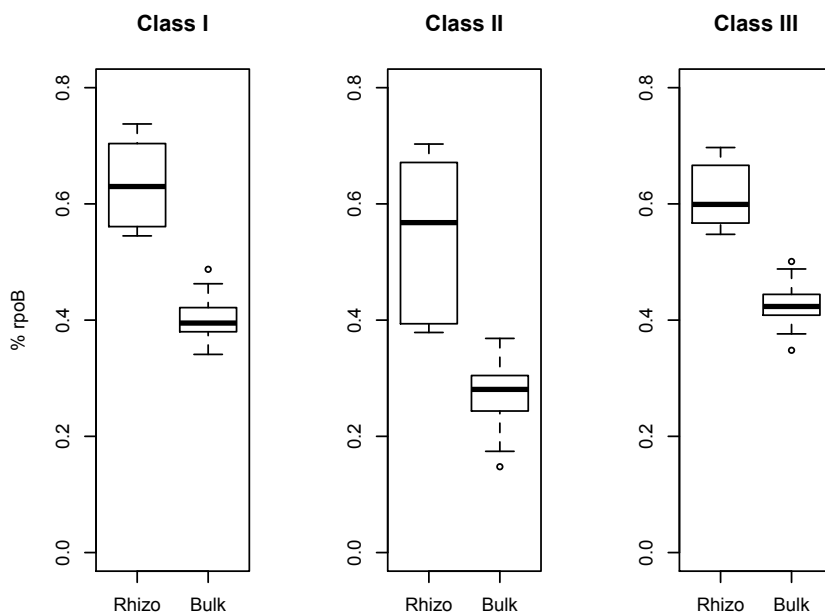
Figure 19. Community abundance of PHA-producing taxa associated with class I, class II and class III *phaC* in bulk soil versus rhizosphere soil metagenomes. Community abundance was estimated by

44

summing the proportional abundance of all genera associated with predicted *phaC* genes in each metagenome sample.

3.5 Primer comparisons and improvement using metagenomic data

Although common PCR primer sets targeting all PHA synthase classes have not yet been developed (Yang et al., 2013), class II *phaC* primers (Solaiman et al., 2000; Zhang et al., 2001) have been developed as well as class I primers (Sheu et al., 2000). Primers from Sheu et al., 2000 have been validated based on successful detection of the *phaC* gene in all 19 known PHA-positive bacteria, and primers from Solaiman et al., 2000 were used to identify phaC1/C2 genes within *P. corrugate*, which was verified to accumulate mcl PHA.

The degenerate primers developed in these two studies have been used by many subsequent studies (Ciesielski et al., 2006; Tzu et al., 2012). However, because these primers have been designed based on culturable species which may account for as little as 1% of microbial species (Amann et al., 1995 and Tai et al., 2014), it is conceivable that they do not reflect the sequence diversity of *phaC* genes that exists in the natural environment.

To examine the extent to which existing primers cover the sequence diversity of natural *phaC* genes and also potentially redesign/optimize them using environmental *phaC* sequences, we compared primers from Sheu et al., 2000 (phaCF1 and phaCR4) and from Solaiman et al., 2000 (I-179L and I-179R) to aligned regions obtained from our metagenomic *phaC* homologs. The primer-associated sequence profiles for metagenomic class I and class II genes were separately calculated and compared to existing primers, from which we identified

similarities and differences that may be useful to optimize primers and improve future detection of novel PHA synthase genes.

3.5.1 Back-translation of metagenomic protein alignments for analysis of primer efficiency

To convert the protein alignment of all metagenomic hits for class I and II back to their corresponding nucleotide alignments, a script was developed and also made publicly available at github (https://github.com/pvah88/BackToDNA). The program takes as input an aligned fasta file of metagenome-derived protein sequences, and replaces each amino acid with the corresponding codon triplet identified in the corresponding DNA sequence read (see Methods). Alignments were visualized using Aliview (Larsson 2014) and Jalview (Waterhouse et al., 2009) and primer region were determined manually by identifying the best sub-alignment matches to Sheu and Solaimon's primers. The percentage similarity between a primer and each corresponding DNA sequence in the aligned primers region was calculated using the compareToPrimer.pl scripts located at https://github.com/doxeylab/metannotateScripts (Petrenko et al., 2015). This script allows for the IUPAC nucleotide ambiguity code (http://www.bioinformatics.org/sms/iupac.html) and therefore accounts for primer degeneracy when available. To visualize the sequence profile for each primer-aligned region, sequence logos were constructed. In addition, to examine the efficiency of each primer in detecting metagenomic PHA synthases, histograms of percentage identity distributions were constructed (**Figure 19, Figure 20**).
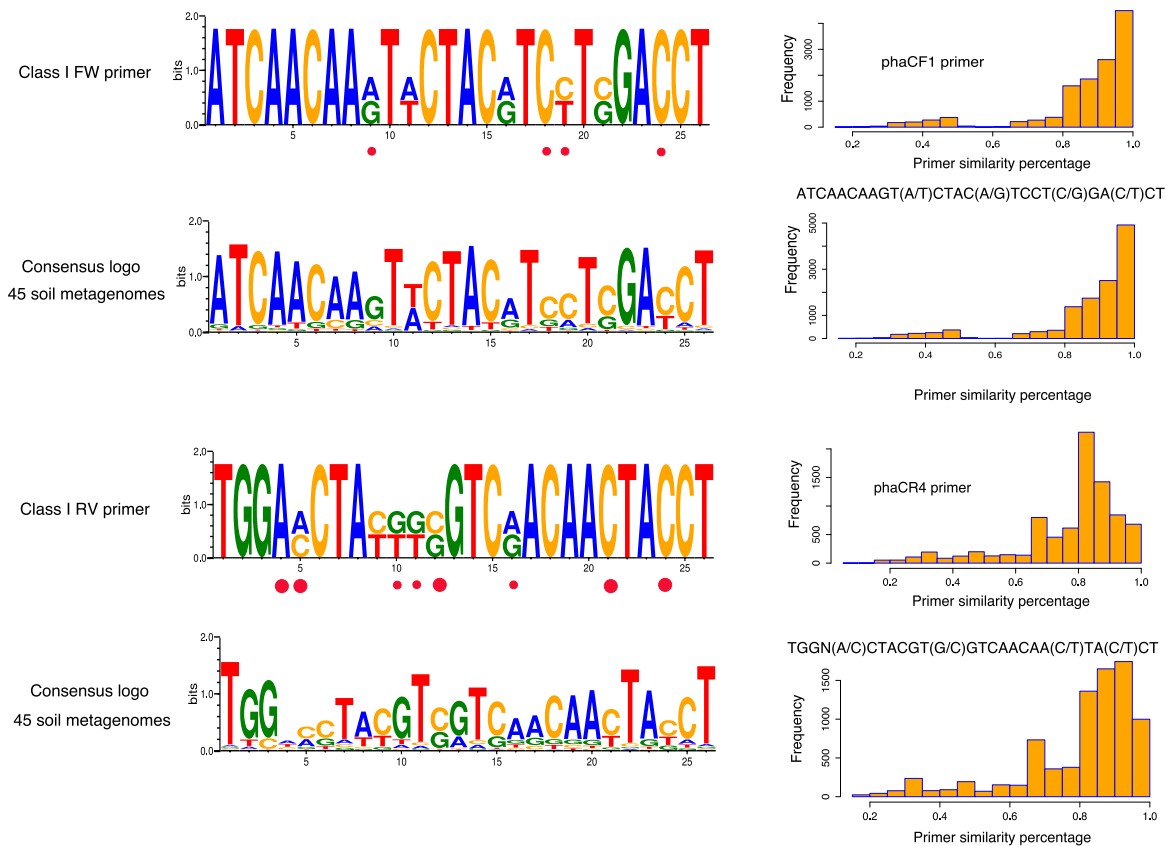
Figure 19. Comparison of metagenomic logos for class I *phaC* genes against aligned primers. Class I forward (phaCF1) and reverse (phaCR4) primers are shown above the metagenomic logos. Positions that have been selected for primer optimization are indicated by red circles (big circles, degenerate positions; small circles, non-degenerate positions). The histograms on the right show the distributions of metagenomic *phaC* sequence identity to the previous primers and optimized primers.

Overall, class I forward and reverse primers from Sheu et al. (2000) were remarkably consistent with the sequence diversity observed in the metagenomic consensus logos (**Figure 19**). Although

47

the logos reveal hidden sequence diversity not present within the primer, the consensus nucleotides from the metagenomic alignment match the primer in almost all positions. In general, class I primers performed better than class II primers from Solaiman et al. (2000) in terms of their congruency with metagenomic sequences (**Figure 20**). The metagenomic class II sequences showed lower similarities to the established class II synthase forward primer (I-179L) with a median percentage identity of 78.6%, and the reverse primer (I-179R) behaved similarity with a median percentage identity of 76.7%.

Despite a high median identity for the majority of metagenomic homologs of class II genes, a surprising number of metagenomic hits aligned very poorly to existing primers. 9.2% and 14.7% of metagenomic hits for class I fw and rv primers were less than 65% identity to the primers and 20.7% and 12.1% of metagenomic hits for class II fw and rv primers were less than 65% identity to the primers. This implies that a sizeable fraction of environmental *phaC* genes will be undetected using existing PCR primers.

To design potential improvements to existing primers, metagenomic *phaC* sequence logos were compared with the primers to identity position-specific inconsistencies in nucleotide values (**Figure 20**). Because primer sequence similarities were already high for class I genes (median forward primer %ID = 86.5%, median reverse primer %ID = 84.6%), there was relatively little room for optimization (**Figure 20, Table 5**). Nevertheless, optimization of the class I reverse primer (phaCR4) through a few substitutions increased the median %ID from 84.6% to 88.5% and resulted in the detection of 4,391 sequence hits above 85% ID versus an original 2,949 (a substantial 48.9% increase). Suggested modification for this primer includes

five degenerate positions and three non-degenerate positions (**Figure 19**). The five degenerate positions are A to any of the four nucleotides (A/G/T/C) at position 4, keeping position 5 and 12 as A/C and C/G respectively, and changing both position 21 and 24 from C to C/T. The three non-degenerate positions are modifications of position 10 and 11 from G/T to G and T respectively, and position 16 from A/G to A.

Because previous class II primers showed weaker congruency with metagenomic *phaC* genes, there was more room to improve their efficiency in detecting environmental sequences. For example, for the forward primer I-L179 from Solaiman et al. (2000) the metagenomic alignment suggested modifications for five degenerate positions and two non-degenerate positions (**Figure 20**). The five suggestions for degenerate positions for class I forward primer were modification of position 19 from C to C/T, modification of position 22 from C to C/G, modification of position 23 from T to any of A/T/C/G, modification of position 25 from C to C/G, and modification of position 28 from C to C/T. The two suggested modifications for the non-degenerate positions were A to G at position 1 and A to G at position 20. Suggestions for modification of the reverse primer I-179R including five degenerate positions and three non-degenerate positions. The five modifications for degenerate positions were: C to G/C, A to any of A/T/C/G, C to C/G, C to C/G and C to C/G at positions 9, 11, 21, 27 and 30, respectively. The three suggestions for non-degenerate positions were A to G, A to C, and A to C at positions 19, 20 and 27 respectively. These modifications to class II degenerate primers resulted in significant improvements in the distribution of percentage similarity of primers (**Figure 20, Table 5**). The median percentage primer identities of metagenomic *phaC* genes increased from

78.6% to 89.3.% (class II forward primer), and 73.6% to 90% (class II reverse primer). For the class II reverse primer, the modified primer is estimated to detect almost seven times as many matches with 85% identity or higher.
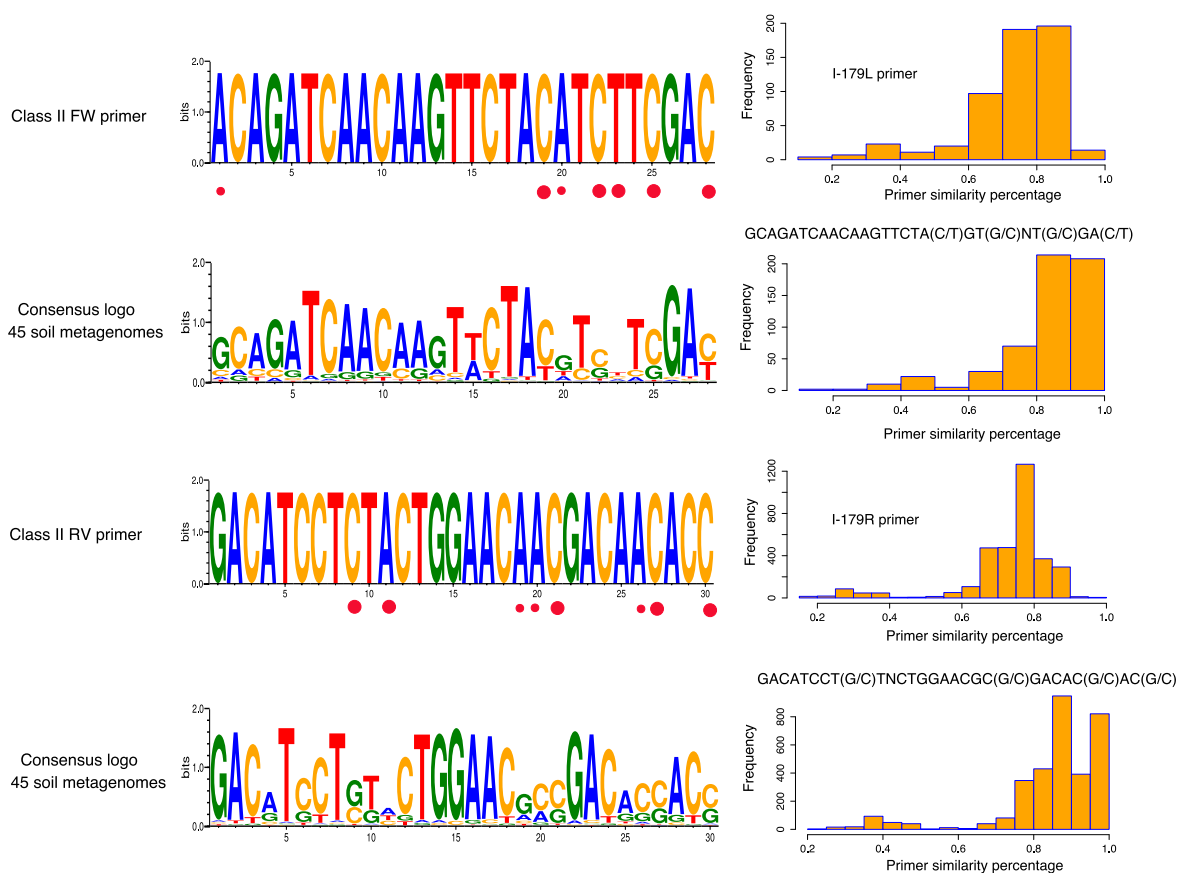
Figure 20. Comparison of metagenomic logos for class II *phaC* genes against aligned primers. Class II forward (I-179L) and reverse (I-179R) primers are shown above the metagenomic logos with big red circles (degenerate positions) and small red circles (non-degenerate positions) indicating positions that have been modified in the optimized primer designs. The histograms on the right show the distributions of metagenomic *phaC* sequence identity to the previous primers and optimized primers.

50

Table 5. Abundance of metagenomic *phaC* sequences matching previous primers versus designed primers

CLASS I	Sheu Forward primers	NEW Class I forward primer
Less than 65%	1150 seqs, 9.2%	1143 seqs, 9.1%
More than 65%	11407 seqs, 90.8%	11414 seqs, 90.9%
More than 85%	8951 seqs, 71.3%	9176 seqs, 73.1%
More than 95%	4491 seqs, 35.8%	4918 seqs, 39.2%
mean	86.5%	87.1%
median	92.3%	92.3%
	Sheu reverse primers	**NEW Class I reverse primer**
Less than 65%	1228 seqs, 14.7%	1113 seqs, 13.4%
More than 65%	7108 seqs, 85.3%	7223 seqs, 86.6%
More than 85%	2949 seqs, 35.4%	4391 seqs, 52.6%
More than 95%	681 seqs, 8.2%	999 seqs, 12%
mean	77.7%	85.2%
median	84,6%	92.3%

CLASS II	Solaiman Forward primers	NEW Class II forward primer
Less than 65%	117 seqs, 20.8%	53 seqs, 9.4%
More than 65%	446 seqs, 79.2%	510 seqs, 90.6%
More than 85%	142 seqs, 25.2%	380 seqs, 67.5%
More than 95%	3 seqs, 5.3%	104 seqs, 18.5%
mean	73.8%	84.1%
median	78.6%	89.3%
	Solaiman reverse primers	**NEW Class II reverse primer**
Less than 65%	398 seqs, 12.1%	237 seqs, 7.2%
More than 65%	2898 seqs, 87.9%	3059 seqs, 92.8%
More than 85%	310 seqs, 9.4%	2161 seqs, 65.6%
More than 95%	5 seqs, 1.5%	821 seqs, 24.9%
mean	73.6%	85.8%
median	76.7%	90%

Chapter 4- Discussion

This thesis reports the first large-scale metagenomic analysis of environmental PHA synthase genes and taxa from diverse soil environments. The analyses reveal new insights with regards to several areas: 1) Taxonomic profiling of metagenomic PHA homologs revealed the identity and abundance of PHA producing taxa in natural environments, which confirmed an important role for PHA production in existing PHA-associated taxa, and also predicted environmentally abundant PHA-producing taxa that are unknown or less well known in this context. 2) Comparison of PHA producer communities between different classes of *phaC* synthases revealed differences (sometimes dramatic) in the taxa that are likely synthesizing mcl and scl PHAs in the environment. 3) Comparison of PHA producer communities between bulk and rhizosphere soils revealed distinct differences in taxonomic structure that are consistent with the bacteria ecological behaviour and further subdivided by soil and plant host type. In addition, an enrichment in the proportional abundance of PHA-producing organisms was detected in rhizosphere metagenomes, which suggests a relatively important role for scl and mcl PHAs in rhizosphere-associated species. 4) Finally, analysis of the sequence novelty of detected metagenomic PHA synthases as well as aligned primer regions suggests that there exists environmental *phaC* genes that are highly unique from reference sequences and existing primers. The metagenomic data can be used to optimize and redesign existing models for phaC genes and primers for future studies, as we have demonstrated through several *in silico* primer modifications.

4.1.1 Identification of known and novel taxa associated with PHA production in soil metagenomes

Using reciprocal HMM-based homology search, which was validated to have high accuracy based on a set of positive controls, we identified 229,070 total metagenomic sequences affiliated with *phaC* genes encoding class I-III PHA synthases. Most taxa predicted to contribute high amount of *phaC* genes in soil metagenomes affiliated are well-established PHA producers such as *Rhizobiaceae* for class I, *Pseudomonas* for class II and *Methylobacteriacea* for class III. However, interestingly, metagenomic data predicted additional taxa (e.g., Class I: *Sphingomonadaceae*, Class II: *Geodermatophilacea*, Class III: *Xanthomonadaceae*), whose *phaC* genes are relatively understudied, as major environmental PHA producers.

The under-appreciation of these taxa is further emphasized by the somewhat counterintuitive result of the sequence novelty of *phaC* sequences detected from selected soil metagenomes. As expected, class I metagenomic hits, especially the well known taxa (i.e., *Rhizobiaceae* and *Bradyrhizobiaceae*), demonstrated high similarity identity to the sequences in NCBI RefSeq database. However, class II and III metagenomic hits from relatively understudied taxa (e.g., Class II: *Geodermatophilaceae*, Class III: *Xanthomonadaceae*) were surprisingly similar to database sequences, and more so than taxa whose *phaC* genes have been widely studied (e.g., Class II: *Pseudomonadaceae*, Class III: *Chromatiaceae*). This suggests that existing reference genomes already encode the potential for PHA synthesis, but that these taxa have not been widely recognized for this function. The prediction of these organisms as dominant environmental producers of PHA places them into the content of PHA metabolism and

53

provides motivation for future work characterizing their PHA pathways and PHA-related end products in more detail. As a similar example from previous literature, despite the existence of several *Thaumarchaeota* genomes, it was not known that *Thaumarchaeota* were cobalamin (B_{12}) producing phyla until they were detected as dominant contributors of B_{12} genes in aquatic metagenomes (Doxey et al., 2015).

4.1.2 Changes in PHA producer communities between environments

The association between PHA production and rhizobacteria has been widely discussed in previous literature, and there has been quite a range of results on the topic (Chen et al., 2010) Although some studies have reported weak or no phenotypic effects observed in *phaC* mutants of rhizobacterial species such as (Povolo et al. 1994; Lodwig et al., 2005), others have found strong support for a link between PHA production and both root colonization and plant-growth promoting activity (Aneja et al., 2004; Lodwig et al., 2005, Fibach-Paldi et al., 2012). For example, coinoculation experiments have shown that wild-type *Sinorhizobium meliloti* strains have a competitive advantage against *phaC* mutants when co-inoculated with their host (alfalfa) plants (Willis and Walker, 1998). In addition, plants inoculated with *phaC* deficient mutants have been found to have reduced rates of nitrogen fixation, root nodulation, and shoot dry weight (Wang et al., 2007). Thus, intracellular stores of PHA may support processes such as nitrogen fixation during periods of carbon starvation, and enable cell division and growth during root colonization (Trainer and Charles, 2006).

One potential reason for the discrepancy in previous experimental results on the effect of PHA production on rhizobacterial fitness and plant interactions is that these interactions may be highly dependent on the particular plant-host system. By directly detecting and characterizing the PHA-producing community in a soil, as done in this thesis using metagenomic analysis, we were able to circumvent these issues and quantify the relative importance of PHA synthesis in the rhizosphere. In doing so, we detected a significant difference in both the taxonomic community structure and community abundance of PHA producers in the rhizosphere versus bulk soil. The community abundance of all taxa contributing class I, class II and class III PHA synthase genes in particular, showed a significant increase in the rhizosphere (**Figure 18 and 19**), which provides further evidence that short chain length and medium chain chain length PHA production provides a competitive advantage to rhizosphere-associated bacteria.

The specific PHA-producing taxa enriched in the rhizosphere tend to be plant growth promoting or plant pathogenic species (e.g., Class I: *Rhizobiaceae, Sphingomonadaceae*; Class II: *Pseudomonadaceace*; Class III: *Xanthomonadaceace*). On the other hand, taxa that are less dependent on host plant carbon sources contributed more to the *phaC* gene pool bulk soils than in the rhizosphere. These species include slow growing species (e.g., *Bradyrhizobiaceae*), photosynthetic species (e.g., *Rhodospirillaceae*), and species known to survive harsh environments (e.g., *Burkhoderiaceae* and *Geodermathophilaceae*).

4.1.3 Sequence and primer novelty of metagenomic *phaC* genes

Based on sequence similarities between metagenomic *phaC* genes and reference sequences from NCBI Refseq, as well as analysis of primer-aligned regions, it can be concluded that there are many novel PHA synthase sequences present in the soil metagenomes that would likely escape detection using traditional (i.e., PCR-based) approaches. Because our metagenomic hits includes novel PHA synthase sequences and cover genomes from both culturable and non-culturable bacteria, an alignment of the metagenomic *phaC* hits provides an excellent guide to optimize existing degenerate *phaC* primers. By comparing class I degenerate primers from Sheu et al., 2000 and class II degenerate primers from Solaiman et al., 2000 to the sequence logos of soil metagenomic *phaC* genes, two new class II (forward and reverse) primers were developed and shown shown to increase efficiency in recognizing metagenomic *phaC* homologs with higher percentage identity. These primers will be used and further tested in future work.

Several computational tools exist for designing primers from multiple sequence alignments (Brodin et al., 2013; Yoon and Leitner, 2015), and previous studies have also used metagenomic sequences to either design or test designed primers. For example, Varaljay et al. (2009) used marine metagenomic data to design primers for the Dimethylsulfoniopropionate Degrading Gene (*dmdA*) and study the diversity and abundance of marine demethylating bacteria. In addition, Coyotzi et al. (2016) used soil metagenomic data to evaluate the efficacy of existing primers for the *nirK* gene. The methodology used here and the application to phaC genes however is different from previous work and thus is a highly novel aspect of this work.

4.1.4 Conclusion and Future Work

Ultimately, analysis of 45 soil metagenomes from a diverse group of soils has provided a broad view of the soil PHA-producer community, how it changes by class (medium chain length, short chain length, PHA synthase class I-III) and environment (bulk versus rhizosphere), as well as the taxonomic and sequence novelty of *phaC* genes in soil environments. From this data, dominant PHA-producing taxa were observed at significantly increased community abundance in the rhizosphere, which provides strong quantitative support of previous hypotheses regarding the role of PHA in plant-microbe interactions. In addition, the metagenomic data has guided the design of improved primer that should increase the coverage of environmental sequences in future studies.

References

Ahemad, M., & Kibret, M. (2014). Mechanisms and applications of plant growth promoting rhizobacteria: Current perspective. *Journal of King Saud University-Science*, *26*(1), 1-20.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, *25*(17), 3389-3402.

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, *59*(1), 143-169.

Amara, A. A., & Moawad, H. (2011). PHA synthases and PHA depolymerases: The enzymes that produce and degrade plastic. *IIUM Engineering Journal*, *12*(4).

Amara, A., Steinbüchel, A., & Rehm, B. (2002). In vivo evolution of the *Aeromonas punctata* polyhydroxyalkanoate (PHA) synthase: isolation and characterization of modified PHA synthases with enhanced activity. *Applied microbiology and biotechnology*, *59*(4-5), 477-482.

Anderson, A. J., & Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiological reviews*, *54*(4), 450-472.

Aneja, K. K., Ashby, R. D., & Solaiman, D. K. (2009). Altered composition of *Ralstonia eutropha* poly (hydroxyalkanoate) through expression of PHA synthase from *Allochromatium vinosum* ATCC 35206. *Biotechnology letters*, *31*(10), 1601-1612.

Aneja, P., Dai, M., Lacorre, D. A., Pillon, B., & Charles, T. C. (2004). Heterologous complementation of the exopolysaccharide synthesis and carbon utilization phenotypes of *Sinorhizobium meliloti* Rm1021 polyhydroxyalkanoate synthesis mutants. *FEMS microbiology letters*, *239*(2), 277-283.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., ... & Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic acids research*, *30*(1), 276-280.

Bernd, H. A. (2003). Polyester synthases: natural catalysts for plastics. *Biochemical Journal*, *376*(1), 15-33.

Bibby, K., & Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental science & technology*, *47*(4), 1945-1951.

Braunegg, G., Lefebvre, G., & Genser, K. F. (1998). Polyhydroxyalkanoates, biopolyesters from renewable resources: physiological and engineering aspects. *Journal of Biotechnology*, *65*(2), 127-161.

Brodin, J., Krishnamoorthy, M., Athreya, G., Fischer, W., Hraber, P., Gleasner, C., ... & Leitner, T. (2013). A multiple-alignment based primer design algorithm for genetically highly variable DNA targets. BMC bioinformatics, 14(1), 1.

Carreto, L., Moore, E., Nobre, M. F., Wait, R., Riley, P. W., Sharp, R. J., & DA COSTA, M. S. (1996). *Rubrobacter xylanophilus* sp. nov., a new thermophilic species isolated from a thermally polluted effluent. *International Journal of Systematic and Evolutionary Microbiology*, *46*(2), 460-465.

Cermola, M., Fedorova, E., Taté, R., Riccio, A., Favre, R., & Patriarca, E. J. (2000). Nodule invasion and symbiosome differentiation during *Rhizobium etli-Phaseolus vulgaris* symbiosis. *Molecular plant-microbe interactions*, *13*(7), 733-741.

Chen, G. Q. (2010). Introduction of Bacterial Plastics PHA, PLA, PBS, PE, PTT, and PPP. In *Plastics from Bacteria* (pp. 1-16). Springer Berlin Heidelberg.

Chen, J. Y., Liu, T., Zheng, Z., Chen, J. C., & Chen, G. Q. (2004). Polyhydroxyalkanoate synthases *phaC1* and *phaC2* from *Pseudomonas stutzeri* 1317 had different substrate specificities. *FEMS microbiology letters*, *234*(2), 231-237.

Chen, M. Y., Wu, S. H., Lin, G. H., Lu, C. P., Lin, Y. T., Chang, W. C., & Tsay, S. S. (2004). *Rubrobacter taiwanensis* sp. nov., a novel thermophilic, radiation-resistant species isolated from hot springs. *International journal of systematic and evolutionary microbiology*, *54*(5), 1849-1855.

Chen, W. M., De Faria, S. M., James, E. K., Elliott, G. N., Lin, K. Y., Chou, J. H., ... & Vandamme, P. (2007). *Burkholderia nodosa* sp. nov., isolated from root nodules of the woody Brazilian legumes *Mimosa bimucronata* and *Mimosa scabrella*. *International Journal of Systematic and Evolutionary Microbiology*, *57*(5), 1055-1059.

Cheng, J., & Charles, T. (2016). Functional metagenomics using *Pseudomonas putida* expands the known diversity of polyhydroxyalkanoate synthases and enables the production of novel polyhydroxyalkanoate copolymers. *bioRxiv*, 042705, doi: 10.1007/s00253-016-7666-6

Ciesielski, S., Cydzik-Kwiatkowska, A., Pokoj, T., & Klimiuk, E. (2006). Molecular detection and diversity of medium-chain-length polyhydroxyalkanoates-producing bacteria enriched from activated sludge. *Journal of applied microbiology*, *101*(1), 190-199.

Cohen-Bazire, G., & Kunisawa, R. (1963). The fine structure of *Rhodospirillum rubrum*. *The Journal of cell biology*, *16*(2), 401-419.

Collins, M. D., Smida, J., Dorsch, M., & Stackebrandt, E. (1988). *Tsukamurella* gen. nov. harboring *Corynebacterium paurometabolum* and *Rhodococcus aurantiacus*. *International Journal of Systematic and Evolutionary Microbiology*, *38*(4), 385-391.

Coyotzi, S., Doxey, A. C., Clark, I. D., Lapen, D. R., Van Cappellen, P., & Neufeld, J. D. (2016). Agricultural soil denitrifiers possess extensive nitrite reductase gene diversity. *Environmental Microbiology*, doi: 10.1111/1462-2920.13643

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, *14*(6), 1188-1190.

Dawes, E. A. (1988). Polyhydroxybutyrate: an intriguing biopolymer. *Bioscience reports*, *8*(6), 537-547.

Do Young Kim, H. W. K., Chung, M. G., & Rhee, Y. H. (2007). Biosynthesis, modification, and biodegradation of bacterial medium-chain-length polyhydroxyalkanoates. *The Journal of Microbiology*, 45(2), 87-97.

Doxey, A. C., Kurtz, D. A., Lynch, M. D., Sauder, L. A., & Neufeld, J. D. (2015). Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. *The ISME journal*, *9*(2), 461-471.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461.

Eppard, M., Krumbein, W. E., Koch, C., Rhiel, E., Staley, J. T., & Stackebrandt, E. (1996). Morphological, physiological, and molecular characterization of actinomycetes isolated from dry soil, rocks, and monument surfaces. *Archives of Microbiology*, *166*(1), 12-22.

Fibach-Paldi, S., Burdman, S., & Okon, Y. (2012). Key physiological properties contributing to rhizosphere adaptation and plant growth promotion abilities of *Azospirillum brasilense*. *FEMS microbiology letters*, *326*(2), 99-108.

Fiedler, S., Steinbüchel, A., & Rehm, B. H. (2000). *PhaG*-mediated synthesis of poly (3-hydroxyalkanoates) consisting of medium-chain-length constituents from nonrelated carbon sources in recombinant *Pseudomonas fragi*. *Applied and environmental microbiology*, *66*(5), 2117-2124.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, doi: 10.1093/nar/gkr367.

Foong, C. P., Lau, N. S., Deguchi, S., Toyofuku, T., Taylor, T. D., Sudesh, K., & Matsui, M. (2014). Whole genome amplification approach reveals novel polyhydroxyalkanoate synthases (*phaCs*) from Japan Trench and Nankai Trough seawater. *BMC microbiology*, *14*(1), 1.

Füchtenbusch, B., & Steinbüchel, A. (1999). Biosynthesis of polyhydroxyalkanoates from low-rank coal liquefaction products by *Pseudomonas oleovorans* and *Rhodococcus ruber*. *Applied microbiology and biotechnology*, *52*(1), 91-95.

Gano-Cohen, K. A., Stokes, P. J., Blanton, M. A., Wendlandt, C. E., Hollowell, A. C., Regus, J. U., ... & Sachs, J. L. (2016). Nonnodulating *Bradyrhizobium* spp. Modulate the Benefits of Legume-*Rhizobium* Mutualism. *Applied and Environmental Microbiology*, *82*(17), 5259-5268.

Glick, B. R. (2012). Plant growth-promoting bacteria: mechanisms and applications. *Scientifica*, *2012, doi:10.6064/2012/963401*.

Godoy, F., Vancanneyt, M., Martinez, M., Steinbüchel, A., Swings, J., & Rehm, B. H. A. (2003). *Sphingopyxischilensis* sp. nov., a chlorophenol-degrading bacterium that accumulates polyhydroxyalkanoate, and transfer of *Sphingomonas alaskensis* to *Sphingopyxis alaskensis comb. nov*. *International journal of systematic and evolutionary microbiology*, *53*(2), 473-477.

Goodchild, D. J., & Bergersen, F. J. (1966). Electron microscopy of the infection and subsequent development of soybean nodule cells. *Journal of bacteriology*, *92*(1), 204-213.

Gordon, M. A. (1964). The genus *Dermatophilus*. *Journal of Bacteriology*, *88*(2), 509-522.

Gorenflo, V., Steinbüchel, A., Marose, S., Rieseberg, M., & Scheper, T. (1999). Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining. *Applied microbiology and biotechnology*, *51*(6), 765-772.

Grunow, A., & Rabenhorst, L. (1868). Flora europaea algarum aquae dulcis et submarinae.

Gtari, M., Essoussi, I., Maaoui, R., Sghaier, H., Boujmil, R., Gury, J., ... & Daffonchio, D. (2012). Contrasted resistance of stone-dwelling *Geodermatophilaceae* species to stresses known to give rise to reactive oxygen species. *FEMS microbiology ecology*, *80*(3), 566-577.

Gyaneshwar, P., Hirsch, A. M., Moulin, L., Chen, W. M., Elliott, G. N., Bontemps, C., ... & Young, J. P. W. (2011). Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. *Molecular plant-microbe interactions*, *24*(11), 1276-1288.

Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. *Nucleic acids research*, *31*(1), 371-373.

Hall, B., Baldwin, J., Rhie, H. G., & Dennis, D. (1998). Cloning of the *Nocardia corallina* polyhydroxyalkanoate synthase gene and production of poly-(3-hydroxybutyrate-co-3-hydroxyhexanoate) and poly-(3-hydroxyvalerate-co-3-hydroxyheptanoate). *Canadian journal of microbiology*, *44*(7), 687-691.

Han, J., Hou, J., Liu, H., Cai, S., Feng, B., Zhou, J., & Xiang, H. (2010). Wide distribution among halophilic *archaea* of a novel polyhydroxyalkanoate synthase subtype with homology to bacterial type III synthases. *Applied and environmental microbiology*, *76*(23), 7811-7819.

Hermann-Krauss, C., Koller, M., Muhr, A., Fasl, H., Stelzer, F., & Braunegg, G. (2013). Archaeal production of polyhydroxyalkanoate (PHA) co-and terpolyesters from biodiesel industry-derived by-products. *Archaea*, *2013, doi:10.1155/2013/129268*.

Hezbri, K., Ghodhbane-Gtari, F., del Carmen Montero-Calasanz, M., Sghaier, H., Rohde, M., Spröer, C., ... & Gtari, M. (2015). *Geodermatophilus aquaeductus* sp. nov., isolated from the ruins of Hadrian's aqueduct. *Antonie van Leeuwenhoek*, *108*(1), 41-50.

Hirsch, A. M., Long, S. R., Bang, M., Haskins, N., & Ausubel, F. M. (1982). Structural studies of alfalfa roots infected with nodulation mutants of *Rhizobium meliloti*. *Journal of bacteriology*, *151*(1), 411-419.

Hoffmann, N., Steinbüchel, A., & Rehm, B. H. (2000a). The *Pseudomonas aeruginosa phaG* gene product is involved in the synthesis of polyhydroxyalkanoic acid consisting of medium-chain-length constituents from non-related carbon sources. *FEMS microbiology letters*, *184*(2), 253-259.

Hoffmann, N., Steinbüchel, A., & Rehm, B. H. A. (2000b). Homologous functional expression of cryptic *phaG* from *Pseudomonas oleovorans* establishes the transacylase-mediated polyhydroxyalkanoate biosynthetic pathway. *Applied microbiology and biotechnology*, *54*(5), 665-670.

Hoppensack, A., Rehm, B. H., & Steinbüchel, A. (1999). Analysis of 4-Phosphopantetheinylation of Polyhydroxybutyrate Synthase from *Ralstonia eutropha*: Generation of β-Alanine Auxotrophic Tn5 Mutants and Cloning of the panD Gene Region. *Journal of bacteriology*, *181*(5), 1429-1435.

Howe, A. C., Jansson, J. K., Malfatti, S. A., Tringe, S. G., Tiedje, J. M., & Brown, C. T. (2014). Tackling soil diversity with the assembly of large, complex metagenomes. Proceedings of the National Academy of Sciences, 111(13), 4904-4909.

Huijberts, G. N., de Rijk, T. C., de Waard, P., & Eggink, G. (1994). 13C nuclear magnetic resonance studies of *Pseudomonas putida* fatty acid metabolic routes involved in poly (3-hydroxyalkanoate) synthesis. *Journal of bacteriology*, *176*(6), 1661-1666.

Huisman, G. W., de Leeuw, O. L. A. V., Eggink, G., & Witholt, B. (1989). Synthesis of poly-3-hydroxyalkanoates is a common feature of fluorescent *pseudomonads*. *Applied and environmental microbiology*, *55*(8), 1949-1954.

Hustede, E., Steinbüchel, A., & Schlegel, H. G. (1992). Cloning of poly (3-hydroxybutyric acid) synthase genes of *Rhodobacter sphaeroides* and *Rhodospirillum rubrum* and heterologous expression in *Alcaligenes eutrophus*. *FEMS microbiology letters*, *93*(3), 285-290.

Ishiguro, E. E., & Fletcher, D. W. (1975). Characterization of *Geodermatophilus* strains isolated from high altitude Mount Everest soils. *Acta Biol Iugosl, Ser B Mikrobiol, 12*, 99-108.

IUPAC Codes. (n.d.). Retrieved November 11, 2016, from http://www.bioinformatics.org/sms/iupac.html

Ivanova, N., Sikorski, J., Jando, M., Munk, C., Lapidus, A., Del Rio, T. G., ... & Chen, F. (2010). Complete genome sequence of *Geodermatophilus obscurus* type strain (G-20 T). *Standards in genomic sciences*, *2*(2), 158.

James, B. W., Mauchline, W. S., Dennis, P. J., Keevil, C. W., & Wait, R. (1999). Poly-3-hydroxybutyrate in *Legionella pneumophila*, an energy source for survival in low-nutrient environments. *Applied and environmental microbiology*, *65*(2), 822-827.

Jendrossek, D., Schirmer, A., & Schlegel, H. G. (1996). Biodegradation of polyhydroxyalkanoic acids. *Applied Microbiology and Biotechnology*, *46*(5-6), 451-463.

Jia, Y., Kappock, T. J., Frick, T., Sinskey, A. J., & Stubbe, J. (2000). Lipases provide a new mechanistic model for polyhydroxybutyrate (PHB) synthases: characterization of the functional residues in *Chromatium vinosum* PHB synthase. *Biochemistry*, *39*(14), 3927-3936.

Jia, Y., Yuan, W., Wodzinska, J., Park, C., Sinskey, A. J., & Stubbe, J. (2001). Mechanistic studies on class I polyhydroxybutyrate (PHB) synthase from *Ralstonia eutropha*: class I and III synthases share a similar catalytic mechanism. *Biochemistry*, *40*(4), 1011-1019.

Jordan, D. C. (1982). NOTES: transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *International Journal of Systematic and Evolutionary Microbiology*, *32*(1), 136-139.

Kadouri, D., Burdman, S., Jurkevitch, E., & Okon, Y. (2002). Identification and isolation of genes involved in poly (β-hydroxybutyrate) biosynthesis in *Azospirillum brasilense* and characterization of a *phbC* mutant. *Applied and environmental microbiology*, *68*(6), 2943-2949.

Kadouri, D., Jurkevitch, E., & Okon, Y. (2003). Involvement of the reserve material poly-β-hydroxybutyrate in *Azospirillum brasilense* stress endurance and root colonization. *Applied and environmental microbiology*, *69*(6), 3244-3250.

Kattar, M. M., Cookson, B. T., Carlson, L. C., Stiglich, S. K., Schwartz, M. A., Nguyen, T. T., ... & Coyle, M. B. (2001). *Tsukamurella strandjordae* sp. nov., a proposed new species causing sepsis. *Journal of clinical microbiology*, *39*(4), 1467-1476.

Khan, A. L., Waqas, M., Kang, S. M., Al-Harrasi, A., Hussain, J., Al-Rawahi, A., ... & Lee, I. J. (2014). Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *Journal of Microbiology*, *52*(8), 689-695.

Kim, Y. B., & Lenz, R. W. (2001). Polyesters from microorganisms. In *Advances in Biochemical Engineering/Biotechnology, 71,* 51-79.

Kudukhashvili, P. G., Gurielidze, M. A., & Pataraya, D. T. (2001). Study of the lytic activities of actinomycetes isolated from different soils in Georgia. *Applied Biochemistry and Microbiology*, *37*(3), 251-252.

Kung, S. S., Chuang, Y. C., Chen, C. H., & Chien, C. C. (2007). Isolation of polyhydroxyalkanoates-producing bacteria using a combination of phenotypic and genotypic approach. *Letters in applied Microbiology*, *44*(4), 364-371.

Langenbach, S., Rehm, B. H., & Steinbüchel, A. (1997). Functional expression of the PHA synthase gene *phaC1* from *Pseudomonas aeruginosa* in *Escherichia coli* results in poly (3-hydroxyalkanoate) synthesis. *FEMS microbiology letters*, *150*(2), 303-309.

LARKIN, J. M., & SHINABARGER, D. L. (1983). Characterization of *Thiothrix nivea*. *International Journal of Systematic and Evolutionary Microbiology*, *33*(4), 841-846.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276-3278.

Léveillé, J. H. (1851). *Organisation et disposition méthodique des especes qui composent le genre Erysiphé*. Martinet.

LIEBERGESELL, M., & STEINBÜCHEL, A. (1992). Cloning and nucleotide sequences of genes relevant for biosynthesis of poly (3hydroxybutyric acid) in *Chromatium vinosum* strain D. *European Journal of Biochemistry*, *209*(1), 135-150.

Liebergesell, M., Hustede, E., Timm, A., Steinbüchel, A., Fuller, R. C., Lenz, R. W., & Schlegel, H. G. (1991). Formation of poly (3-hydroxyalkanoates) by phototrophic and chemolithotrophic bacteria. *Archives of microbiology*, *155*(5), 415-421.

Liebergesell, M., Sonomoto, K., Madkour, M., Mayer, F., & Steinbüchel, A. (1994). Purification and characterization of the poly (hydroxyalkanoic acid) synthase from *Chromatium vinosum* and localization of the enzyme at the surface of poly (hydroxyalkanoic acid) granules. *European Journal of Biochemistry*, *226*(1), 71-80.

Liu, L., Jiang, C. Y., Liu, X. Y., Wu, J. F., Han, J. G., & Liu, S. J. (2007). Plant–microbe association for rhizoremediation of chloronitroaromatic pollutants with Comamonas sp. strain CNB-1. *Environmental Microbiology*, *9*(2), 465-473.

Liu, W. T., Hanada, S., Marsh, T. L., Kamagata, Y., & Nakamura, K. (2002). *Kineosphaera limosa* gen. nov., sp. nov., a novel Gram-positive polyhydroxyalkanoate-accumulating coccus isolated from activated sludge. *International journal of systematic and evolutionary microbiology*, *52*(5), 1845-1849.

Lodwig, E. M., Hosie, A. H., Bourdès, A., Findlay, K., Allaway, D., Karunakaran, R., ... & Poole, P. S. (2003). Amino-acid cycling drives nitrogen fixation in the legume–Rhizobium symbiosis. *Nature*, *422*(6933), 722-726.

Lodwig, E. M., Leonard, M., Marroqui, S., Wheeler, T. R., Findlay, K., Downie, J. A., & Poole, P. S. (2005). Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids. *Molecular plant-microbe interactions*, *18*(1), 67-74.

Lodwig, E., & Poole, P. (2003). Metabolism of Rhizobium bacteroids. *Critical reviews in plant sciences*, *22*(1), 37-78.

Lu, Q., Han, J., Zhou, L., Zhou, J., & Xiang, H. (2008). Genetic and biochemical characterization of the poly (3-hydroxybutyrate-co-3-hydroxyvalerate) synthase in *Haloferax mediterranei*. *Journal of bacteriology*, *190*(12), 4173-4180.

Luedemann, G. M. (1968). *Geodermatophilus*, a new genus of the *Dermatophilaceae* (Actinomycetales). *Journal of bacteriology*, *96*(5), 1848-1858.

Madigan, M. T., Martinko, J. M., Parker, J., Fernández, E. G., Fernández, C. R., & Pérez, M. S. (2004). Brock Biología de los microorganismos.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., ... & Lanczycki, C. J. (2014). CDD: NCBI's conserved domain database. *Nucleic acids research*, gku1221, doi: 10.1093/nar/gku1221.

Masalha, J., Kosegarten, H., Elmaci, Ö., & Mengel, K. (2000). The central role of microbial activity for iron acquisition in maize and sunflower. *Biology and Fertility of Soils*, *30*(5-6), 433-439.

McCool, G. J., & Cannon, M. C. (2001). *PhaC* and *PhaR* are required for polyhydroxyalkanoic acid synthase activity in *Bacillus megaterium*. *Journal of Bacteriology*, *183*(14), 4235-4243.

Montero-Calasanz, M. D. C., Göker, M., Rohde, M., Spröer, C., Schumann, P., Mayilraj, S., ... & Klenk, H. P. (2014). Description of *Geodermatophilus amargosae* sp. nov., to accommodate the not validly named *Geodermatophilus obscurus* subsp. *amargosae* (Luedemann, 1968). *Current microbiology*, *68*(3), 365.

Munk, A. C., Lapidus, A., Lucas, S., Nolan, M., Tice, H., Cheng, J. F., ... & Huntemann, M. (2011). Complete genome sequence of *Tsukamurella paurometabola* type strain (no. 33 T). *Standards in genomic sciences*, *4*(3), 342.

Nam, S. W., Chun, J., Kim, S., Kim, W., Zakrzewska-Czerwinska, J., & Goodfellow, M. (2003). *Tsukamurella spumae* sp. nov., a novel actinomycete associated with foaming in activated sludge plants. *Systematic and applied microbiology*, *26*(3), 367-375.

Nam, S. W., Kim, W., Chun, J., & Goodfellow, M. (2004). *Tsukamurella pseudospumae* sp. nov., a novel actinomycete isolated from activated sludge foam. *International journal of systematic and evolutionary microbiology*, *54*(4), 1209-1212.

Olson, J. B., Harmody, D. K., Bej, A. K., & McCarthy, P. J. (2007). *Tsukamurella spongiae* sp. nov., a novel actinomycete isolated from a deep-water marine sponge. *International journal of systematic and evolutionary microbiology*, *57*(7), 1478-1481.

Ostle, A. G., & Holt, J. G. (1982). Nile blue A as a fluorescent stain for poly-beta-hydroxybutyrate. *Applied and Environmental Microbiology*, *44*(1), 238-241.

Park, S. W., Kim, S. M., Park, S. T., & Kim, Y. M. (2009). *Tsukamurella carboxydivorans* sp. nov., a carbon monoxide-oxidizing actinomycete. *International journal of systematic and evolutionary microbiology*, *59*(6), 1541-1544.

Petrenko, P., Lobb, B., Kurtz, D. A., Neufeld, J. D., & Doxey, A. C. (2015). MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. *BMC biology*, *13*(1), 92.

Poindexter, J. S. (1981). The caulobacters: ubiquitous unusual bacteria. *Microbiological Reviews*, *45*(1), 123.

Poli, A., Di Donato, P., Abbamondi, G. R., & Nicolaus, B. (2011). Synthesis, production, and biotechnological applications of exopolysaccharides and polyhydroxyalkanoates by archaea. *Archaea*, *2011,* http://dx.doi.org/10.1155/2011/693253.

Povolo, S., Tombolini, R., Morea, A., Anderson, A. J., Casella, S., & Nuti, M. P. (1994). Isolation and characterization of mutants of *Rhizobium meliloti* unable to synthesize poly-β-hydroxybutyrate. *Canadian journal of microbiology*, *40*(10), 823-829.

Qi, Q., & Rehm, B. H. (2001). Polyhydroxybutyrate biosynthesis in *Caulobacter crescentus*: molecular characterization of the polyhydroxybutyrate synthase. *Microbiology*, *147*(12), 3353-3358.

Qi, Q., Rehm, B. H., & Steinbüchel, A. (1997). Synthesis of poly (3-hydroxyalkanoates) in *Escherichia coli* expressing the PHA synthase gene *phaC2* from *Pseudomonas aeruginosa*: comparison of *PhaC1* and *PhaC2*. *FEMS microbiology letters*, *157*(1), 155-162.

Qi, Q., Steinbüchel, A., & Rehm, B. H. (1998). Metabolic routing towards polyhydroxyalkanoic acid synthesis in recombinant *Escherichia coli* (*fadR*): inhibition of fatty acid β-oxidation by acrylic acid. *FEMS microbiology letters*, *167*(1), 89-94.

Qi, Q., Steinbüchel, A., & Rehm, B. H. A. (2000). In vitro synthesis of poly (3-hydroxydecanoate): purification and enzymatic characterization of type II polyhydroxyalkanoate synthases *PhaC1* and *PhaC2* from *Pseudomonas aeruginosa*. *Applied microbiology and biotechnology*, *54*(1), 37-43.

Quelas, J. I., Mongiardini, E. J., Pérez-Giménez, J., Parisi, G., & Lodeiro, A. R. (2013). Analysis of two polyhydroxyalkanoate synthases in *Bradyrhizobium japonicum* USDA 110. *Journal of bacteriology*, *195*(14), 3145-3155.

Raaijmakers, J. M., Bitter, W., Punte, H. L., Bakker, P. A., Weisbeek, P. J., & Schippers, B. (1994). Siderophore receptor PupA as a marker to monitor wild-type *Pseudomonas putida* WCS358 in natural environments. *Applied and environmental microbiology*, *60*(4), 1184-1190.

Rabenhorst, L. 1864–1868. *Flora europaea algarum aquae dulcis et submarinae*, *3*.

Rainey, F. A., Ray, K., Ferreira, M., Gatz, B. Z., Nobre, M. F., Bagaley, D., ... & Small, A. M. (2005). Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Applied and Environmental Microbiology*, *71*(9), 5225-5235.

Ratcliff, W. C., Kadam, S. V., & Denison, R. F. (2008). Poly-3-hydroxybutyrate (PHB) supports survival and reproduction in starving rhizobia. *FEMS Microbiology Ecology*, *65*(3), 391-399.

Rehm B.A.H. (2003). Polyester synthases: natural catalysts for plastics. *Biochemical Journal*, *376*(1), 15-33.

Rehm, B. H. (2007). Biogenesis of microbial polyhydroxyalkanoate granules: a platform technology for the production of tailor-made bioparticles. *Current issues in molecular biology*, *9*(1), 41.

Reslewic, S., Zhou, S., Place, M., Zhang, Y., Briska, A., Goldstein, S., ... & Lapidus, A. (2005). Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*. *Applied and environmental microbiology*, *71*(9), 5511-5522.

Rey, D., Fraisse, P., Riegel, P., Piemont, Y., & Lang, J. M. (1997). [*Tsukamurella* infections. Review of the literature apropos of a case]. *Pathologie-biologie*, *45*(1), 60-65.

Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic acids research*, *38*(20), e191-e191.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics*, *16*(6), 276-277.

Ruiz, J. A., López, N. I., Fernández, R. O., & Méndez, B. S. (2001). Polyhydroxyalkanoate Degradation Is Associated with Nucleotide Accumulation and Enhances Stress Resistance and Survival of Pseudomonas oleovorans in Natural Water Microcosms. *Applied and environmental microbiology*, *67*(1), 225-230.

Salisbury, F. B. (1994). The role of plant hormones. *Plant-environment interactions. Marcel Dekker, New York*, 39-81.

Sandhya, V., Ali, S. Z., Grover, M., Reddy, G., & Venkateswarlu, B. (2010). Effect of plant growth promoting *Pseudomonas* spp. on compatible solutes, antioxidant status and plant growth of maize under drought stress. *Plant growth regulation*, *62*(1), 21-30.

Schalk, I. J., Hennard, C., Dugave, C., Poole, K., Abdallah, M. A., & Pattus, F. (2001). Iron-free pyoverdin binds to its outer membrane receptor FpvA in *Pseudomonas aeruginosa*: a new mechanism for membrane iron transport. *Molecular microbiology*, *39*(2), 351-361.

Schallmey, M., Ly, A., Wang, C., Meglei, G., Voget, S., Streit, W. R., ... & Charles, T. C. (2011). Harvesting of novel polyhydroxyalkanoate (PHA) synthase encoding genes from a soil metagenome library using phenotypic screening. *FEMS microbiology letters*, *321*(2), 150-156.

Schlegel, H. G., Lafferty, R., & Krauss, I. (1970). The isolation of mutants not accumulating poly-β-hydroxybutyric acid. *Archiv für Mikrobiologie*, *71*(3), 283-294.

Schwartz, M. A., Tabet, S. R., Collier, A. C., Wallis, C. K., Carlson, L. C., Nguyen, T. T., ... & Coyle, M. B. (2002). Central Venous Catheter–Related Bacteremia Due to *Tsukamurella* Species in the Immunocompromised Host: A Case Series and Review of the Literature. *Clinical Infectious Diseases*, *35*(7), e72-e77.

Seong, C. N., Kim, Y. S., Baik, K. S., Choi, S. K., Kim, M. B., Kim, S. B., & Goodfellow, M. (2003). *Tsukamurella sunchonensis* sp. nov., a bacterium associated with foam in activated sludge. *The Journal of Microbiology*, *41*(2), 83-88.

Sharma, A., & Johri, B. N. (2003). Growth promoting influence of siderophore-producing *Pseudomonas* strains GRP3A and PRS 9 in maize (Zea mays L.) under iron limiting conditions. *Microbiological research*, *158*(3), 243-248.

Sheu, D. S., Wang, Y. T., & Lee, C. Y. (2000). Rapid detection of polyhydroxyalkanoate-accumulating bacteria isolated from the environment by colony PCR. *Microbiology*, *146*(8), 2019-2025.

Solaiman, D. K., Ashby, R. D., & Foglia, T. A. (2000). Rapid and specific identification of medium-chain-length polyhydroxyalkanoate synthase gene by polymerase chain reaction. *Applied microbiology and biotechnology*, *53*(6), 690-694.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., ... & Böhm, C. (2012). The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environmental microbiology*, *14*(12), 3122-3145.

Spiekermann, P., Rehm, B. H., Kalscheuer, R., Baumeister, D., & Steinbüchel, A. (1999). A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. *Archives of Microbiology*, *171*(2), 73-80.

Stackebrandt, E., Kroppenstedt, R. M., & Fowler, V. J. (1983). A phylogenetic analysis of the family Dermatophilaceae. *Microbiology*, *129*(6), 1831-1838.

Stackebrandt, E., Rainey, F. A., & Ward-Rainey, N. L. (1997). Proposal for a new hierarchic classification system, *Actinobacteria classis* nov. *International Journal of Systematic and Evolutionary Microbiology*, *47*(2), 479-491.

Steinbüchel, A., & Füchtenbusch, B. (1998). Bacterial and other biological systems for polyester production. *Trends in biotechnology*, *16*(10), 419-427.

Steinbüchel, A., & Schlegel, H. G. (1991). Physiology and molecular genetics of poly (β-hydroxyalkanoic acid) synthesis in *Alcaligenes eutrophus*. *Molecular microbiology*, *5*(3), 535-542.

Stowers, M. D. (1985). Carbon metabolism in *Rhizobium* species. *Annual Reviews in Microbiology*, *39*(1), 89-108.

Sturgeon, A., Stull, J. W., Costa, M. C., & Weese, J. S. (2013). Metagenomic analysis of the canine oral cavity as revealed by high-throughput pyrosequencing of the 16S rRNA gene. *Veterinary microbiology*, *162*(2), 891-898.

Sudesh, K., Fukui, T., & Doi, Y. (1998). Genetic Analysis of *Comamonas acidovorans* Polyhydroxyalkanoate Synthase and Factors Affecting the Incorporation of 4-Hydroxybutyrate Monomer. *Applied and environmental microbiology*, *64*(9), 3437-3443.

Sukweenadhi, J., Kim, Y. J., Kang, C. H., Farh, M. E. A., Nguyen, N. L., Hoang, V. A., ... & Yang, D. C. (2015). *Sphingomonas panaciterrae* sp. nov., a plant growth-promoting bacterium isolated from soil of a ginseng field. *Archives of microbiology*, *197*(8), 973-981.

Suzuki, K. I., Collins, M. D., Iijima, E., & Komagata, K. (1988). Chemotaxonomic characterization of a radiotolerant bacterium, *Arthrobacter radiotolerans*: description of Rubrobacter radiotolerans gen. nov., comb. nov. *FEMS microbiology letters*, *52*(1-2), 33-39.

Tai, Y. T., Foong, C. P., Najimudin, N., & Sudesh, K. (2016). Discovery of a new polyhydroxyalkanoate synthase from limestone soil through metagenomic approach. *Journal of bioscience and bioengineering*, *121*(4), 355-364.

Tal, S., & Okon, Y. (1985). Production of the reserve material poly-β-hydroxybutyrate and its function in *Azospirillum brasilense* Cd. *Canadian journal of microbiology*, *31*(7), 608-613.

Tao, T. S., Yue, Y. Y., Chen, W. X., & Chen, W. F. (2004). Proposal of *Nakamurella* gen. nov. as a substitute for the bacterial genus *Microsphaera* Yoshimi et al. 1996 and *Nakamurellaceae* fam. nov. as a substitute for the illegitimate bacterial family *Microsphaeraceae* Rainey et al. 1997. *International journal of systematic and evolutionary microbiology*, *54*(3), 999-1000.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., & Tolstoy, I. (2013). RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic acids research*, doi: 10.1093/nar/gkt1274.

Timm, A. R. N. U. L. F., & Steinbüchel, A. (1990). Formation of polyesters consisting of medium-chain-length 3-hydroxyalkanoic acids from gluconate by *Pseudomonas aeruginosa* and other fluorescent *pseudomonads. Applied and environmental microbiology*, *56*(11), 3360-3367.

TIMM, A., & STEINBÜCHEL, A. (1992). Cloning and molecular analysis of the poly (3-hydroxyalkanoic acid) gene locus of *Pseudomonas aeruginosa* PAO1. *European Journal of Biochemistry*, *209*(1), 15-30.

Timm, A., Wiese, S., & Steinbüchel, A. (1994). A general method for identification of polyhydroxyalkanoic acid synthase genes from pseudomonads belonging to the rRNA homology group I. *Applied microbiology and biotechnology*, *40*(5), 669-675.

Tombolini, R., & Nuti, M. P. (1989). Poly (β-hydroxyalkanoate) biosynthesis and accumulation by different *Rhizobium* species. *FEMS microbiology letters*, *60*(3), 299-304.

Tombolini, R., Povolo, S., Buson, A., Squartini, A., & Nuti, M. P. (1995). Poly-β-hydroxybutyrate (PHB) biosynthetic genes in *Rhizobium meliloti* 41. *Microbiology*, *141*(10), 2553-2559.

Trainer, M. A., & Charles, T. C. (2006). The role of PHB metabolism in the symbiosis of rhizobia with legumes. *Applied microbiology and biotechnology*, *71*(4), 377-386.

Tsuge, T., Hyakutake, M., & Mizuno, K. (2015). Class IV polyhydroxyalkanoate (PHA) synthases and PHA-producing *Bacillus. Applied microbiology and biotechnology*, *99*(15), 6231-6240.

Tsukamura, M. (1971). Proposal of a new genus, *Gordona*, for slightly acid-fast organisms occurring in sputa of patients with pulmonary disease and in soil. *Microbiology*, *68*(1), 15-26.

Tzu, H. Y., & Semblante, G. U. (2012). Detection of polyhydroxyalkanoate-accumulating bacteria from domestic wastewater treatment plant using highly sensitive PCR primers. *Journal of microbiology and biotechnology*, *22*(8), 1141-1147.

Valentin, H. E., & Steinbüchel, A. (1993). Cloning and characterization of the *Methylobacterium extorquens* polyhydroxyalkanoic-acid-synthase structural gene. *Applied microbiology and biotechnology*, *39*(3), 309-317.

Varaljay, V. A., Howard, E. C., Sun, S., & Moran, M. A. (2010). Deep sequencing of a dimethylsulfoniopropionate-degrading gene (dmdA) by using PCR primer pairs designed on the basis of marine metagenomic data. *Applied and environmental microbiology*, *76*(2), 609-617.

Wang, C., Meek, D. J., Panchal, P., Boruvka, N., Archibald, F. S., Driscoll, B. T., & Charles, T. C. (2006). Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and environmental microbiology*, *72*(1), 384-391.

Wang, C., Saldanha, M., Sheng, X., Shelswell, K. J., Walsh, K. T., Sobral, B. W., & Charles, T. C. (2007). Roles of poly-3-hydroxybutyrate (PHB) and glycogen in symbiosis of *Sinorhizobium meliloti* with *Medicago* sp. *Microbiology*, *153*(2), 388-398.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191.

Weiner, R. M. (1997). Biopolymers from marine prokaryotes. *Trends in biotechnology*, *15*(10), 390-394.

Willis, L. B., & Walker, G. C. (1998). The phbC (poly-β-hydroxybutyrate synthase) gene of *Rhizobium* (*Sinorhizobium*) *meliloti* and characterization of *phbC* mutants. *Canadian journal of microbiology*, *44*(6), 554-564.

Winogradsky, S. (1888). Beitrage zur Morphologie und Physiologie der Bakterien. I. Zur Morphologie and Physiologie der Schwefelbakterien. A. Felix, Leipzig. 1949 Revision. Republished as Contribution a la morphologie et physiologie des sulfobacteries. *Microbiologie du sol. Massonet Cie, Paris*, 83-126.

Wittenborn, E. C., Jost, M., Wei, Y., Stubbe, J., & Drennan, C. L. (2016). Structure of the catalytic domain of the class I polyhydroxybutyrate synthase from *Cupriavidus necator*. *Journal of Biological Chemistry*, *291*(48), 25264-25277.

Yassin, A. F., Rainey, F. A., Brzezinka, H., Burghardt, J., Lee, H. J., & Schaal, K. P. (1995). *Tsukamurella inchonensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *45*(3), 522-527.

Yassin, A. F., Rainey, F. A., Brzezinka, H., Burghardt, J., Rifai, M., Seifert, P., ... & Schaal, K. P. (1996). *Tsukamurella pulmonis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *46*(2), 429-436.

Yee, L. N., Chuah, J. A., Chong, M. L., Phang, L. Y., Raha, A. R., Sudesh, K., & Hassan, M. A. (2012). Molecular characterisation of *phaCAB* from *Comamonas* sp. EB172 for functional expression in *Escherichia coli* JM109. *Microbiological research*, *167*(9), 550-557.

Yoon, H., & Leitner, T. (2015). PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics, 31(9), 1472-1474.

Yoshimi, Y., Hiraishi, A., & Nakamura, K. (1996). Isolation and characterization of *Microsphaera multipartita* gen. nov., sp. nov., a polysaccharide-accumulating Gram-positive bacterium from activated sludge. *International Journal of Systematic and Evolutionary Microbiology*, *46*(2), 519-525.

Zhalnina, K. V., Dias, R., Leonard, M. T., de Quadros, P. D., Camargo, F. A., Drew, J. C., ... & Triplett, E. W. (2014). Genome sequence of *Candidatus Nitrososphaera evergladensis* from group I. 1b enriched from everglades soil reveals novel genomic features of the ammonia-oxidizing *archaea*. *PloS one*, *9*(7), e101648.

Zhang, S., Kolvek, S., Goodwin, S., & Lenz, R. W. (2004). Poly (hydroxyalkanoic acid) Biosynthesis in *Ectothiorhodospira s haposhnikovii*: Characterization and Reactivity of a Type III PHA Synthase. *Biomacromolecules*, *5*(1), 40-48.

Zinn, M., Witholt, B., & Egli, T. (2001). Occurrence, synthesis and medical application of bacterial polyhydroxyalkanoate. *Advanced drug delivery reviews*, *53*(1), 5-21.