Functional metagenomics using *Pseudomonas putida* expands the known diversity of polyhydroxyalkanoate synthases and enables the production of novel polyhydroxyalkanoate copolymers

Jiujun Cheng and Trevor C. Charles*

Department of Biology and Centre for Bioengineering and Biotechnology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

*Corresponding author:
Dr. Trevor C. Charles
Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.
E-mail: tcharles@uwaterloo.ca

Abstract

Bacterially produced biodegradable polyhydroxyalkanoates with versatile properties can be achieved using different PHA synthase enzymes. This work aims to expand the diversity of known PHA synthases via functional metagenomics, and demonstrates the use of these novel enzymes in PHA production. Complementation of a PHA synthesis deficient *Pseudomonas putida* strain with a soil metagenomic cosmid library retrieved 27 clones expressing either Class I, Class II or unclassified PHA synthases, and many did not have close sequence matches to known PHA synthases. The composition of PHA produced by these clones was dependent on both the supplied growth substrates and the nature of the PHA synthase, with various combinations of SCL- and MCL-PHA. These data demonstrate the ability to isolate diverse genes for PHA synthesis by functional metagenomics, and their use for the production of a variety of PHA polymer and copolymer mixtures.

Keywords:

Functional metagenomics; soil metagenomic library; polyhydroxyalkanoate (PHA); bioplastics; polyhydroxyalkanoate synthase; *Pseudomonas putida*; *Sinorhizobium meliloti*

Introduction

Polyhydroxyalkanoates (PHAs) are natural polyesters biosynthesized by a variety of bacteria under unbalanced growth conditions. They serve as reserves of carbon and reducing power and aid in survival during starvation or stress conditions (Verlinden et al. 2007). These biodegradable and environmentally friendly polymers can be used as alternative materials to conventional petrochemical-based plastics. PHAs are classified into short-chain-length (SCL, C3-C5) and medium-chain-length (MCL, ≥C6), and copolymers (SCL + MCL) based on the number of carbon atoms per monomer. MCL-PHA is generally more useful than SCL-PHA due to it being less brittle and more flexible. To produce PHA with versatile properties cost-effectively, strategies have involved mining new PHA synthase enzymes (PhaC), engineering PhaC proteins and modifying the metabolic pathways of the production strains (Keshavarz and Roy 2010; Schallmey et al. 2011; Cheema et al. 2012; Park et al. 2012; Tripathi et al. 2013; Meng et al. 2014).

PHA molecules consist of over 150 possible constituent monomers of hydroxyl fatty acids (Meng et al. 2014) (Steinbüchel and Lütke-Eversloh 2003). PHA synthase enzymes catalyze the joining of the hydroxyl group of one monomer with the carboxyl group of another by an ester bond to form PHA polymers. The monomeric composition of PHA is determined primarily by PhaC, although the available carbon source and metabolic pathways also influence the properties of PHA (Verlinden et al. 2007). PhaC proteins are grouped into four classes based on amino acid sequence, substrate specificity and subunit composition (Rehm 2003). Class I and II PhaC consist of one subunit (PhaC), but those in Class III and IV are composed of two subunits (PhaC+PhaE and PhaC+PhaR, respectively). Class I and IV PhaCs synthesize SCL-PHA whereas Class II polymerize MCL-PHA. Class III PhaCs can synthesize both SCL and MCL monomers.

54 The key Class I PHA synthesis pathway genes include *phaC* (or *phbC*), *phaA* (or *phbA*) and
55 *phaB* (or *phbB*). The three genes are often, but not always, clustered in a single operon. The
56 acetoacetyl-CoA reductase (EC 2.3.1.9) encoded by the *phaA* and β-ketothiolase (EC 1.1.1.36)
57 encoded by the *phaB* convert acetyl-CoA to (R)-3-hydroxybutyryl-CoA (3HB-CoA) through
58 acetoacetyl-CoA. The Class-I PhaC (EC 2.3.1.-) then polymerizes 3HB-CoA into
59 polyhydroxybutyrate (PHB) (Rehm 2003). The Class II *pha* cluster is well conserved in
60 *Pseudomonas* and consists of two PHA synthase genes (*phaC1* and *phaC2*) flanking a PHA
61 depolymerase gene (*phaZ*), and *phaD* encoding a transcriptional activator of *pha* genes. Phasin-
62 encoding *phaF* and *phaI* gene are transcribed divergently to other *pha* genes. MCL monomers ((R)-
63 3-hydroxylacyl-CoA) are derived from β-oxidation of fatty acids or fatty acid *de novo* synthesis from
64 unrelated carbon sources (Tortajada et al. 2013).
65 Traditional culture-based strategies for obtaining new biocatalysts are limited by the inability
66 to cultivate the majority of environmental microbes. While sequence-based metagenomics can
67 identify genes homologous to those present in available sequence databases, it is difficult to reliably
68 predict the function of truly new genes through homology-based analysis. In contrast, functional
69 metagenomics involves the construction of gene libraries from microbial community genomic DNA
70 and functional screening for novel enzymes of interest or potential industrial applications (Simon
71 and Daniel 2011). Functional metagenomics has the potential to identify truly novel sequences for a
72 given function.
73 In previous work, functional metagenomics was used to isolate new Class I PhaC from soil
74 metagenomic clones by Nile red staining and phenotypic screening in *α-Proteobacteria*
75 *Sinorhizobium meliloti* (Schallmey et al. 2011). In another study, *phaC* genes encoding both Class I
76 and II PhaC proteins were PCR amplified from oil-contaminated soil library clones (Cheema et al.

2012). One of the isolated genes produced PHA copolymer when expressed in *Pseudomonas putida*. Moreover, partial *phaC* genes were also obtained from metagenomic DNA via direct PCR amplification (Foong et al. 2014; Pärnänen et al. 2015; Tai et al. 2015). Building on these previous studies, we constructed a PHA$^-$ strain of *P. putida* KT2440 to use as a surrogate host and functionally identified *phaC* genes after screening millions of agricultural wheat soil metagenomic clones. A total of 27 PHA$^+$ clones were obtained. Accumulation and monomer composition of PHA directed by seven of these clones were further examined.

Methods

Bacterial strains, plasmids, cosmids and growth conditions

Bacterial strains, plasmids and cosmids are listed in Table 1. *E. coli* strains were grown at 37°C in LB (Lennox) medium (1% tryptone (w/v), 0.5% yeast extract (w/v), and 0.5% NaCl (w/v), pH 7). *Pseudomonas* strains were grown at 30°C in LB or 0.1N M63 minimal medium (Escapa et al. 2011) supplemented with 0.5% sodium octanoate (w/v), 0.5% nonanoic acid (v/v) or 1% gluconic acid (w/v). *S. meliloti* was grown in LB or YM medium (Schallmey et al. 2011). Antibiotics were used at the following concentrations: streptomycin, 200 µg/ml for *S. meliloti* and 100 µg/ml for *E. coli*; kanamycin, 100 µg/ml for *Pseudomonas* and 50 µg/ml for *E. coli*; neomycin, 200 µg/ml; rifampicin, 100 µg/ml; gentamicin, 10 µg/ml for *E. coli* and 100 for *P. putida*; and tetracycline 20 µg/ml for *E. coli* or 40 µg/ml for *P. putida*.

Construction of PHA$^-$ strain PpUW2

DNA oligonucleotides are listed in Table 2. A Rif^R spontaneous mutant PpUW1 of *P. putida* KT2440 was generated by plating a culture of strain KT2440 on a LB Rif plate, followed by single

colony purification. To construct the PHA deficient strain PpUW2 (*phaC1ZC2*) of *P. putida* PpUW1, a 943-bp DNA fragment containing 5′-*phaC1* region (766 bp) was PCR amplified using *P. putida* KT2440 genomic DNA as a template and primer pair JC161-JC162, digested with HindIII-BamHI and then cloned into the same sites in pJQ200-SK (Quandt and Hynes 1993), yielding plasmid pJC63. Another 876-bp DNA fragment containing the 3′-*phaC2* gene (120 bp) was PCR amplified using primers JC163-JC164, digested with BamHI-SalI and then cloned into the same sites in pJC63 to obtain plasmid pJC69. An omega-Km cassette was obtained from pHP45Ω-Km (Fellay et al. 1987) by BamHI digestion and then inserted into the same site in pJC69 to obtain pJC71. Plasmid pJC71 was then conjugated into *P. putida* PpUW1 in a triparental mating using helper plasmid pRK600. Single cross-over recombination of pJC71 into the *P. putida* chromosome was selected with Rif and Gm. A double cross-over event was achieved by growing a single Rif^R Gm^R colony overnight in LB, making serial dilutions and then spreading on LB Km with 5% sucrose. The resulting PHA^- strain PpUW2 was verified by examining Gm sensitivity (lost plasmid backbone) and absence of PHA production in LB supplemented with 0.5% octanoate (w/v), and was further confirmed by PCR amplification analysis.

Phenotypic screening for *phaC* genes from metagenomic library clones

Construction of the metagenomic DNA library (11AW) of agricultural wheat soil was described previously (Cheng et al. 2014). The 11AW library contains 9×10^6 clones hosted in *E. coli* HB101. The pooled library clones (0.5 ml of 300 ml stock) were conjugated *en masse* into *P. putida* PpUW2 (PHA^-) with the helper plasmid pRK600. Mating mixture was diluted serially with 0.85% NaCl and ~20,000 transconjugants recovered on each LB Tc plate (15 mm × 150 mm) supplemented with 0.5% Na octanoate (w/v). The plates were incubated at 30°C for 24 h and then kept at 22°C for

2-6 days. Potential PHA^+ clones of opaque white colour were streak purified on the LB octanoate plates, and verified on 0.1N M63 minimal medium plates (Escapa et al. 2011) supplemented with 0.5% octanoate (w/v) and Nile red (0.5 μg/ml). PHA^+ cosmids were transferred from *P. putida* to *E. coli* DH5α(Rif^R), mobilized by pRK600. The cosmid DNA was isolated from *E. coli* strains using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), digested by EcoRI-HindIII-BamHI (Thermo Scientific), and then resolved on 1% TAE agarose gels. Cosmids with distinct electrophoretic patterns were conjugated back to *P. putida* PpUW2 (PHA^-) to confirm their PHA^+ phenotype.

Complementation of *S. meliloti* (*phbC*)

11AW cosmid DNA encoding Class I and II *phaC* genes was introduced into *S. meliloti* Rm11476 (Schallmey et al. 2011) via triparental conjugation. Transconjugants were selected on LB SmNmTc plates, and then streaked on YM plates containing Nile red (Schallmey et al. 2011)for visualizing PHB production.

DNA sequencing and bioinformatics

KOD Xtreme DNA polymerase (Novagen) was used for all PCR. Primers are listed in Table 2. PCR reactions consisted of one cycle of 94°C for 5 min, 30 cycles of 94°C for 30 s, 53°C for 30 s and 68°C for 30 s, and final extension at 68°C for 10 min. The internal regions of *phaC* genes were PCR amplified from PHA^+ cosmids using primer phaCF1 and phaCR4 (Sheu et al. 2000). When no PCR product of correct size was obtained, a semi-nested PCR was performed with the primers phaCF2 and phaCR4 as described previously (Sheu et al. 2000). PCR products were resolved on 2% agarose gels, isolated from the gels using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic),

and then cloned into pJET1.2 vector (Thermo Scientific). Sequences of the cloned partial *phaC* genes were obtained using pJET1.2 sequencing primers (Table 2).

End sequences of PHA^+ cosmid DNA were obtained by Sanger sequencing using universal M13F and M13R primers. For high throughput sequencing, 68 cosmids of 11AW PHA^+ clones and 28 cosmids from other research projects were grouped into 24 pools. Tagmentation of pooled cosmid DNA, PCR amplification and clean-up, and library normalization were performed with the Nextera XT DNA library and index kits (Illumina), according to the supplier's recommendation. The library was sequenced using MiSeq 500-cycle version 2 reagent (Illumina). DNA sequences were assembled using SPAdes Genome Assembler 3.5.0 (BaseSpace, Illumina). Cosmid clones were identified based on the available Sanger end sequences (Lam et al. 2014) and ORFs were annotated with MetaGeneMark (Zhu et al. 2010). Predicted protein sequences were analyzed by BLAST against the non-redundant protein databases (NCBI). Multiple-sequence alignment was performed using MUSCLE (Edgar 2004). Phylogenetic analysis was performed using MEGA6 (Tamura et al. 2013). Origin of cloned metagenomic DNA was predicted by PhyloPythia (Patil et al. 2012).

Cloning of *phaC1*$_{Pp}$ gene of *P. putida* KT2440

To construct a *phaC* expression cosmid for a positive control, the *phaC1*$_{Pp}$ gene (*Pp_5003*) of *P. putida* KT2440 was PCR amplified using primers JC159-JC160 and cloned into the HindIII and BamHI sites in the broad-host-range vector pRK7813 (Cheng et al. 2014), to obtain construct pJC67 (P*lac*::*phaC1*$_{Pp}$). A stop codon (TGA) for terminating the translation of *lacZα* upstream of the *phaC1*$_{Pp}$, and a ribosome-binding site (AGGAG) were incorporated into the primer JC159 (Table 2).

Subcloning *phaC* genes of 11AW metagenomic clone 16

168 The Class II *phaC1*$_{16}$ gene of 11AW clone 16 was PCR amplified with oligos JC301 and
169 JC302, and cloned into the HindIII-EcoRI sites in pRK7813 to obtain pJC123. The *phaC2*$_{16}$ was
170 amplified using primers JC337 and JC334, and inserted into the EcoRI site in pJC123, yielding
171 pJC157. The orientation of cloned *phaC2*$_{16}$ was verified by restriction enzyme mapping. The
172 *phaC2*$_{16}$ gene was also obtained by PCR using oligos JC333 and JC334, and cloned into the HindIII-
173 EcoRI sites in pRK7813 to yield pJC159. A stop codon (TGA) for terminating the translation of
174 *lacZα* gene and ribosome-binding sites were added in primers JC333 and JC337. The cloned *phaC*
175 genes were verified by DNA sequencing.
176
177 **Estimation of PHA production by fluorescent spectrometry**
178 A modification of the previously described method (Wu et al. 2003) was used. *S. meliloti*
179 cells (2.5%, v/v) were added to YM medium. All cultures were incubated at 30^{o}C and 200 rpm for
180 48-72 h. The OD_{600} values of 200 µl cultures were measured in 96-well microtiter plates with the
181 Multiskan Spectrum spectrophotometer (Thermo Labsystems). Samples (180 µl) were stained with
182 20 µl of Nile red (2.5 µg/ml) in the dark for 1 h. Fluorescent intensity was measured at excitation
183 (485 nm) and emission (595 nm) with the FilterMax F5 Multi-mode microplate reader (Molecular
184 Devices). PHA content was calculated based on the equation (fluorescent intensity/OD_{600}×0.9).
185
186 **PHA production, extraction and characterization**
187 *P. putida* PpUW1 (PHA^+) or PpUW2 (PHA^-) carrying PHA^+ clones were grown overnight in
188 LB with or without Tc, washed once with 0.85% NaCl, and then subcultured (2%, v/v) in the 0.1N
189 M63 medium with or without Tc, supplemented with 0.5% Na octanoate, 0.5% nonanoic acid or 1%
190 gluconic acid. The cultures were grown at 30°C and at 200 rpm for 48 hrs. Cells were collected by

centrifugation at 20°C and at 9000 × *g* for 15 min, washed once with deionized water. Cell dry weight (CDW) was obtained after drying the cells at 95°C for 24 hr. For PHA methanolysis, cell pellet (~15 mg) was suspended in 2 ml of chloroform and 2 ml of methanol with 15% H_2SO_4 (v/v), and incubated at 100°C for 5 h. After the reaction mixture was cooled down to 20°C, 1 ml of deionized water was added, and vortexed vigorously for 1 min. The chloroform phase was passed through a cotton plug in a Pasteur glass pipette to remove any cell debris. Methanolyzed sample (1 µl) was analyzed with GC-MS (an Agilent 7975B GC equipped with Agilent 5975B inert XL EI/CI MSD and an HP-5MS capillary column). Oven temperature was run at initial 50°C for 5 min with a ramp of 20°C/min to 280°C, and then held for 10 min. The flow rate of Helium carrier gas was 1.2 ml/min. Methylated PHA monomers were identified using the Agilent enhanced MSD chemstation (E.02.01.117). PHA standards were kindly provided by Dr. Bruce A. Ramsay (Polyferm Canada).

Results

Isolation of metagenomic clones for PHA production

To employ *P. putida* as a surrogate host to isolate 11AW (agricultural wheat soil) metagenomic library clones encoding functional PhaC, it was first necessary to construct a PHA synthesis mutant. Expression of the contiguous three genes encoding PHA synthase $PhaC1_{Pp}$ (Pp_5003), PHA depolymerase $PhaZ_{Pp}$ (Pp_5004) and PHA synthase $PhaC2_{Pp}$ (Pp_5005) in *P. putida* KT2440 Rif^R derivative PpUW1 was disrupted by deletion of 3437-bp comprised of the 3′ region (911 bp) of $phaC1_{Pp}$ (1677 bp), $phaZ_{Pp}$ and the 5′ region (1560 bp) of $phaC2_{Pp}$ (1680 bp) and replacement with an omega-Km kanamycin resistance insert, resulting in the PHA^- strain PpUW2. Transfer of the negative control cosmid vector pJC8 to *P. putida* PpUW2 did not result in detectable PHA production. However, introduction of pJC67 (P*lac*::$phaC1_{Pp}$) *in trans* restored the PHA^+

phenotype (Fig. 1). These data suggested that the strain PpUW2 could be used for screening of PhaC-encoding metagenomic clones. The 11AW library clones (Cheng et al. 2014) were transferred to *P. putida* PpUW2 via en masse triparental conjugation. Following selection of *P. putida* PpUW2 transconjugants (~4 million) on LB Km Tc plates supplemented with 0.5% Na octanoate, we obtained 72 clones that exhibited greater opacity than the PpUW2 PHA^- recipient strain. The PHA-producing phenotype of those clones was verified by visualizing the fluorescence of Nile red-stained PHA in 0.1N M63 minimal medium (Escapa et al. 2011) supplemented with Na octanoate as the sole carbon source (Fig. 1). Restriction digest of DNA from the 72 clones demonstrated 68 distinct restriction patterns, suggesting the presence of a broad diversity of DNA origin in those PHA^+ clones.

Identification of *phaC* genes encoding PHA synthases

Internal regions (~500 bp) of the *phaC* genes of 18 distinct cosmid clones were initially obtained by PCR amplification using degenerate primers PhaCF1 and PhaCR4 (Sheu et al. 2000). For additional 4 clones, *phaC* fragments ($PhaC_3$, $PhaC_7$, $PhaC_{10}$, and $PhaC_{15}$) of ~400 bp were generated by nested-PCR using primer pair PhaCF2-PhaCR4 as described previously (Sheu et al. 2000). BLASTP analysis of the cloned gene products indicated that 13 PhaC proteins could be grouped into Class I while the other 9 were categorized as Class II PHA synthases.

To identify all *pha* genes on the isolated clones, DNA sequences were obtained by high throughput sequencing, and additional *pha* gene loci were identified in partially and fully assembled clones (Table 3; Supplementary 1). The identified PHA synthases could be classified into 3 groups based on amino acid sequences (Fig. 2). Class I *phaC* genes were annotated in 17 clones. The *phaC* and *phaAB* genes were adjacently located in 12 of these clones, whereas the *phaB* gene was located distantly downstream of *phaCA* genes in clone 25 and clone P1N3 (Supplementary 1). A *phaR* gene

was located immediately downstream of *phaCB* genes in partially assembled clone P11N2. The *phaC* and *phaB* genes flanked a *phaZ* gene in the partially assembled clone P2N8. The metagenomic DNA in these Class I clones was predicted to originate from *Gemmatimonas*, *α-Proteobacteria* (*Sphingomonadaceae*), *β-Proteobacteria* (*Leptothrix, Rubrivivax, Janthinobacterium*) and γ-*Proteobacteria* (*Xanthomonadaceae*) (Table 3).

Class II PHA genes are commonly clustered with *phaZ* flanked by two PHA synthase encoding genes, *phaC1* and *phaC2*. Nine clones carried Class II *phaC* genes (Table 3). The *pha* gene locus (*phaC1*$_1$-*phaZ*-*phaC2*$_1$-*phaD*-*phaF*-*phaI*) in clone 1 was similar to the canonical locus in *Pseudomonas* including *P. putida* KT2440 (de Eugenio et al. 2010), except the presence of *orf7* encoding a PHA granule associated protein (Pfam09650) (Fig. 3A). Clone 2 and clone 15 were identical to the clone 1, except that they contained an insertion of the 17,013-bp transposon Tn4652 at 13,740 nt and 16,362 nt respectively (Supplementary 2). The Tn4652 duplicated target sequences were AACTC in clone 2 and TAGGA in clone 15. The transposon insertions in these clones are in opposite orientation. The same Tn4652 is present in *P. putida* KT2440 genome (3,366,550 - 3,383,562 nt, GenBank: AE015451), located distant from the *pha* gene operon. These insertions likely occurred following introduction of the cosmid clones into *P. putida*.

Class II clones 5, 16, 20 and 50 had no genes encoding *phaD* or *phaFI* homologs. A *phaJ* gene encoding R-specific enoyl-CoA hydratase was identified in those clones (Fig. 3A; Supplementary 1). The *phaJ* gene was located immediately upstream of the *phaC* genes in clones 16, 20 and 50, but downstream of *phaC* in clone 5. In addition, a gene encoding PHA depolymerase (PhaZ) was located downstream of the *phaC* genes in clones 16, 20 and 50. ORFs downstream of the *phaZ* genes in clones 16 and 50 encoded proteins (PhaC2$_{16}$ and PhaC2$_{50}$) homologous to unclassified

PHA synthases (Fig. 2; Supplementary 1). The proteins shared conserved regions in PhaC enzymes, but were ~30% shorter in N-terminal regions than in those of Class I/II PhaC (Supplementary 2).

Two ORFs encoding putative PhaC and PhaE proteins were identified in partially assembled clone P2N10 (Table 3; Supplementary 1). BLASTP showed the P2N10 PhaC protein best matched to the Class III PhaC [GenBank: WP_002708071] (43% identity) from γ-*Proteobacteria Thiothrix nivea* (Table 3). The available C-terminal sequence of P2N10 PhaE (234 amino acids) only exhibited 26% identity to the Class III PhaE [GenBank: WP_002708072] of *T. nivea*. In addition, the P2N10 PhaC protein was only 29% identical to the Class IV PhaC [GenBank: AAD05620] of *Bacillus megaterium*. Phylogenetic analysis suggested that P2N10 PHA synthase was distant from Class III/IV PhaC and might represent a new subclass of PHA synthases (Fig. 2).

Multiple alignment of amino acid sequences of each of the identified 11AW PhaC proteins showed the conserved catalytic triad (C287, D445 and H473 in $PhaC1_{16}$), a tryptophan essential for dimerization (W391 in $PhaC1_{16}$), and the lipase box GXCXGG (Jia et al. 2000) except that Ala replaced the first Gly residue in $PhaC_{20}$ and second Gly in clone P8N7 (Fig. 3B; Supplementary 3). The serine residue (S229 in $PhaC1_{16}$) required for PhaC activity (Hoppensack et al. 1999) was also conserved in all 11AW PhaC proteins. We chose the PhaC proteins in the Class I clones 14, 18 and 25, and the Class II clones 1, 16 and 20, for further study. We examined the nature of PHA produced using different carbon sources that result in substrate production either through fatty acid synthesis or β-oxidation.

Class II clone 1 synthesizes MCL PHA

The cloned metagenomic DNA in clone 1 contained 42,747 bp (Table 3) [GenBank: KT944254] and most likely originated from *γ-Proteobacteria Pseudomonas*. The $PhaC1_1$, PhaZ,

282 PhaC2$_1$ and PhaD proteins [GenBank: ALV86243, ALV86281, ALV86274 and ALV86275] were
283 82%, 92%, 73% and 78% identical to the corresponding orthologs (PhaC1, PhaZ, PhaC2 and PhaD)
284 of *P. putida* KT2440. Clone 1 PhaC2$_1$ was phylogenetically related to the PHA synthase 2
285 [GenBank: BAA36202] of *Pseudomonas* sp. 61-3 (Fig. 2). In contrast to other Class II PhaC, the
286 conserved amino acid at position 129 of PhaC2$_1$ was Asp rather than Glu (Fig. 3B). It has been
287 previously demonstrated that substitution of the conserved Glu with Asp in Class II PhaC improves
288 PHA yield with an increase of 3-hydroxybutyrate monomer (Matsumoto et al. 2005).
289 PHA synthesis in clone 1 was likely contributed solely by PhaC2$_1$ because the 5'-region of
290 *phaC1$_1$* was absent in the cloned DNA. Expression of the functional *phaZ* and *phaC2$_1$* might be
291 driven by the promoters upstream of the individual genes, as occurs in *P. putida* KT2440 (de
292 Eugenio et al. 2010). *P. putida* PpUW2 (PHA$^-$) carrying clone 1 synthesized 3-hydroxyhexanote
293 (3HHx, C6) and 3-hydroxyoctanoate (3HO, C8) copolymer of ~95% C8 monomer when grown with
294 octanoate, similar to the PHA produced by wild-type *P. putida* PpUW1 (Table 4). When PpUW2
295 (clone 1) was grown with nonanoic acid, 3-hydroxynonanoate (3HN, C9) and 3-hydroxyheptanoate
296 (3HP, C7) were incorporated into the PHA with greater 3HP than that synthesized by wild type
297 PpUW1 (Table 4). These data suggest that those monomers were derived from β-oxidation of
298 octanoate and nonanoic acid.
299 When gluconic acid was used as the sole carbon source, 3HO and 3-hydroxyldecanoate
300 (3HD, C10) copolymer was produced in PpUW2 (PHA$^-$) carrying clone 1, very similar to the
301 production of the parental strain PpUW1 (PHA$^+$) (Table 4). The monomer composition of PHA was
302 similar in both strains, but total PHA accumulation was ~4 fold higher in gluconic acid-grown
303 PpUW2 with clone 1 (Table 4). These results suggest that Class II PhaC1$_2$ was able to synthesize
304 MCL (C6-C10) PHA.

305

306 **Class II clone 16 synthesizes SCL-MCL copolymer**

307 Clone 16 contains 33,900-bp metagenomic DNA [GenBank: KT944263] probably originated
308 from *β-Proteobacteria Methylibium* (Table 3). A *phaJ* gene (*phaJ*$_{16}$) is predicted to encode a R-
309 specific enoyl-CoA hydratase [GenBank: ALV86416] (Fig. 3A), which converts the β-oxidation
310 intermediate 2-enoyl-CoA to (R)-3-hydroxyacyl-CoA used as PhaC substrate (Fukui et al. 1998).
311 The PhaJ$_{16}$ was phylogenetically related to the PhaJ4a$_{Re}$ and PhaJ4b$_{Re}$ of *Ralstonia eutropha*
312 (Kawashima et al. 2012), and PhaJ4$_{Pp}$ of *P. putida* (Sato et al. 2011) (Supplementary 4A). The
313 amino acid residues Asp39 and His44 conserved at the active site of the dehydratases were identified
314 in PhaJ$_{16}$ (Supplementary 3B), except that the Ser residue was replaced by Pro71 in PhaJ$_{16}$, same as
315 in the PhaJ orthologs of *R. eutropha* and *P. putida* PhaJ4.

316 The three ORFs downstream of the *phaJ*$_{16}$ gene encoded PhaC1$_{16}$, PhaZ and PhaC2$_{16}$ (Fig.
317 3A). PhaC1$_{16}$ of Class II PHA synthase most closely matched the *Pseudomonas* sp. F15 PhaC1
318 [GenBank: WP_021487788], exhibiting 48% amino acid identity. PhaC1$_{16}$ was also 42.5% and
319 43.4% identical to the PhaC1 and PhaC2 proteins of *P. putida* KT2440. Thr318 and Val478 in the
320 PhaC1$_{16}$ might favour the 3HB incorporation into PHA as occurred in the engineered PhaC proteins
321 (Takase et al. 2003; Chen et al. 2014). Another putative PHA synthase encoded by the *phaC2*$_{16}$ gene
322 downstream of the *phaC1*$_{16}$*phaZ* genes was homologous to unclassified PHA synthases (Fig. 2).

323 The quantity of PHA accumulated in *P. putida* UW2 carrying clone 16 was comparable to that
324 of parental PpUW1 when grown with octanoate, nonanoic acid or gluconic acid (Table 4). In
325 contrast to the MCL PHA (C6-C8) present in octanoate-grown PpUW2 (clone 1) or wild-type
326 PpUW1, about 20% 3HB was detected in the PHA (C4-C6-C8) accumulated in PpUW2 carrying
327 clone 16 under the same conditions (Table 4). When PpUW2 (clone 16) was grown with nonanoic

328 acid (C9), 3-hydroxyheptanoate (3HP, C7), 3-hydroxyvalerate (3HV, C5) and 3HB (C4) were
329 polymerized, which was in contrast to the 3HP-3HN PHA in nonanoic acid-grown PpUW with
330 PhaC1 or wild type PpUW1. Additionally, equal amounts of C6 and C8 monomers were detected in
331 the PHA from gluconic acid-grown PpUW2 with clone 16. However, PHA with C8 and C10
332 monomers was accumulated in PpUW2 (clone 1) and PpUW1 under the same growth conditions.
333 These data implied that the PhaC1$_{16}$ and/or PhaC2$_{16}$ prefer substrates with shorter carbon chains than
334 the PhaC2$_1$ in clone 1 and both PhaC proteins of *P. putida* KT2440.
335 To further elucidate the activity of PhaC1$_{16}$ and PhaC2$_{16}$, the genes were cloned
336 downstream of the constitutively active P*lac* promoter. When *P. putida* PpUW2 carrying pJC123
337 (P*lac*::*phaC1$_{16}$*) was grown with octanoate, SCL-MCL copolymer was of similar composition of C4-
338 C6-C8 monomers to that synthesized by clone 16 (Table 4). However, PHA was not detected in
339 strain PpUW2 carrying pJC159 (P*lac*::*phaC2$_{16}$*) under the same growth conditions. Coexpression of
340 *phaC1$_{16}$* and *phaC2$_{16}$* in pJC157 (*Plac*::*phaC1$_{16}$-phaC2 $_{16}$*) in octanoate-grown PpUW2 resulted in
341 synthesis of PHA with the similar monomer composition as those in the polymers produced in
342 PpUW2 carrying clone 16 or pJC123 (P*lac*::*phaC1$_{16}$*) (Table 4). These data indicated that only the
343 PhaC1$_{16}$ was involved in SCL-MCL PHA biosynthesis in clone 16.
344 The *pha* gene locus of clone 16 was very similar to that of clone 50 (Supplementary 1), but
345 the DNA in clone 50 probably originated from γ-*Proteobacteria Hydrocarboniphata* (Table 3). Both
346 the Class II PhaC1 and unclassified PhaC2 in clones 16 and 50 shared 84% identity respectively.
347 Expression of the *pha* genes was probably driven by a promoter upstream of *phaJ* genes.
348
349 **Class II clone 20 synthesizes MCL or SCL-MCL PHA**

PHA$^+$ clone 20 contained 30,964-bp metagenomic DNA [GenBank: KT944266], which probably originated from *Salinibacter* within the *Bacteroidetes* (Table 3). The cloned DNA contained only a single *phaC*$_{20}$ gene (Fig. 3A). A putative fatty acid CoA ligase/ synthetase [GenBank: ALV86493] (ORF9; Supplementary 1) might be involved in degradation of fatty acids to acyl-CoA (the first step of β-oxidation cycle). The acyl dehydratase PhaJ$_{20}$ [GenBank: ALV86492] was 56% identical to the PhaJ$_{16}$. Conserved amino acid residues Asp41 and His46 at the active sites of the acyl dehydratases were present (Supplementary 4B) except that the Ser residue was replaced by Pro73 in the PhaJ$_{20}$ protein. The *phaJ*$_{20}$, *phaC*$_{20}$ and *phaZ* genes in clone 20 were similar to those of the corresponding genes in clones 16 and 50 (Fig. 3A; Supplementary 1). In addition, the *pha* gene cluster of clone 20 was similar to that in *Burkholderia* sp. 383 (1,576,365 -1,581,635 nt) [GenBank: NC_007511], but a gene encoding a phasin protein is located between the *phaC* and *phaZ* genes in the *Burkholderia* genome.

The single PhaC$_{20}$ [GenBank: ALV86493] best matched to its ortholog from *Burkholderia lata* at 59% amino acid sequence identity [GenBank: WP_011354994]. PhaC$_{20}$ was 46-53% identical to the PhaC1$_{16}$, PhaC2 and PhaC$_5$ as well as PhaC1 and PhaC2 of *P. putida* KT2440. Presence of amino acid residues Thr353 and Val516 in PhaC$_{20}$ (Fig. 3B) might favour substrates with fewer carbons than those preferred by clone 1 and *P. putida* PhaC proteins, as previously demonstrated with engineered Class II PHA synthases (Takase et al. 2003; Chen et al. 2014).

Clone 20 was able to synthesize 3HHx-3HO copolymer (C6-C8) with octanoate (Table 4), as occurred in PpUW2 (clone 1) and *P. putida* wild type. However, the proportion of C6 monomer increased 4-35 fold in the PHA isolated from PpUW2 (clone 20), compared to that in PpUW2 (clone 1) and wild-type PpUW1 (Table 4). When nonanoic acid was supplied, PpUW2 (clone 20) produced PHA with 8% 3-hydroxyvalerate (3HV), which was absent in the PHA synthesized by clone 1 and

373 PpUW1 (Table 4). In contrast to the SCL-MCL PHA produced by clone 16, 3HB was not detected in
374 the nonanoic acid or octanoate-grown PpUW2 with clone 20.

375 In contrast to the absence of detectable C6 monomer in the PHA produced in PpUW2 with
376 clone 1 or *P. putida* wild type, 3HHx (~10%) was incorporated into the MCL PHA (C6-C8-C10) in
377 PpUW2 (clone 20) (Table 4) when gluconic acid was used as the sole carbon source. About half the
378 quantity of monomers was C10 in the PHA accumulated in PpUW2 (clone 20) grown with gluconic
379 acid. These results suggested that the carbon chain lengths of favourite monomers of $PhaC_{20}$ were
380 between those of $PhaC1_{16}$ and $PhaC2_{1}$.

381

382 **Class I clone 14 synthesizes SCL-MCL PHA**

383 The DNA in clone 14 probably originated from *β-Proteobacteria Methylibium* (Table 3). The
384 $PhaC_{14}$, PhaA and PhaB [GenBank: ALV86397, ALV86398 and ALV86399] were 65%, 76% and
385 74% identical to the PhbC1 (A16_1437), PhbA (A16_1438) and PhbB1 (A16_1439) of *C. necator*
386 H16, respectively. In addition, $PhaC_{14}$ was 48% identical to the PhbC protein (Smc002960) of *S.*
387 *meliloti* Rm1021. The best match of $PhaC_{14}$ was the PhaC protein of *Burkholderia* JOSHI_001 at
388 74% (Table 3). The *phaC*$_{14}$*AB* locus of clone 14 is similar to that of clones 8, 21 and 23
389 (Supplementary 1). Expression of the *phaC*$_{14}$, *phaA* and *phaB* genes in clone 14 was most likely
390 driven by a promoter upstream of the *orf9* encoding a hypothetical protein.

391 When PpUW2 (clone 14) was grown with nonanoic acid, 7% 3HB was incorporated into
392 3HB-3HV copolymer (Table 4). 3HB was the primary monomer of PHA in octanoate- or gluconic
393 acid-grown PpUW2. Both C6 and C8 monomer were present in the PHA with gluconic acid as
394 carbon source. However, C8 monomer was absent in the PHA when octanoate was supplied (Table
395 4). *P. putida* PpUW2 carrying clone 14 produced the highest quantity of PHA among all the PHA^{+}

clones (Table 4). These results indicated that $PhaC_{14}$ could synthesize SCL-MCL copolymer though 3HB was the dominant monomer.

Class I clone 18 synthesizes SCL and SCL-MCL PHA

Clone 18 contained 37,818-bp metagenomic DNA [GenBank: KT944264], which probably originated from *β-Proteobacteria Rubrivirax* (Table 3). The annotated *phaC*$_{18}$, *phaA* and *phaB* were organized in one operon (Fig. 3A), as also occurred in clones 27 and 51 (Supplementary 1). $PhaC_{18}$ [GenBank: ALV86462] shared 100% and 99% identity to $PhaC_{27}$ [GenBank: ALV86602] and $PhaC_{51}$ [GenBank: ALV86651] respectively. The $PhaC_{18}$ protein best matched to its homolog in *Azohydromonas australica* (Table 3), and was 71%, 49% and 61% identical to the $PhaC_{14}$, $PhbC_{Sm}$ and $PhbC1_{Re}$ respectively.

SCL-MCL PHA (C4-C6-C8) was produced when PpUW2 (clone 18) was grown with octanoate (Table 4), as occurred in PpUW2 with clone 16. However, the fraction of C6-C8 monomers was only 3% in the PHA from clone 18, in contrast to the ~83% in the copolymer from clone 16. The yield of PHA was ~7 fold less than that produced by clone 14, but C8 monomer was absent (Table 4). PHA of 3HB-3HV monomers was accumulated in nonanoic acid-grown PpUW2 (clone 18). The ratio of 3HB to 3HV increased 5 fold compared to that in PpUW2 (clone 14) under the same growth conditions (Table 4). When PpUW2 carrying clone 18 was grown on gluconic acid, only PHB was produced. The yield of polymer increased 5 fold compared to that in octanoate-grown cells, and also the second highest quantity of PHA produced among all strains grown on gluconic acid. These results implied that $PhaC_{18}$ could synthesize SCL- and SCL-MCL PHA dependent of the available carbon source.

Class I clone 19 synthesizes SCL and SCL-MCL PHA

Clone 19 contained 25,085-bp of DNA that might originate from *β-Protoebacteria Leptothrix* (Table 3) [GenBank: KT944265]. Three ORFs encoded $PhaC_{19}$, PhaA and PhaB for PHA biosynthesis (Fig. 3A). $PhaC_{19}$ shared 99% identity with $PhaC_6$ and $PhaC_3$ (Fig. 2). The best match for $PhaC_{19}$ is the ortholog of *A. austrilica* at 73% (Table 3). The *phaC*$_{19}$*AB* genes in clone 19 probably consisted of one operon, and the promoter was located upstream of the *phaC*$_{19}$.

Clone 19 was able to synthesize SCL-MCL PHA (C4-C6-C8) with octanoate but only PHB (C4) in gluconic acid-grown PpUW2 (Table 4), as occurred in PpUW2 carrying clone 18. However ~5 fold more PHA was synthesized by clone 19 than was produced by clone 18 in octanoate medium (Table 4). In contrast, the quantity of PHA was lower in gluconic acid-grown PpUW2 with clone 19 than it was in the same host carrying clone 18. Moreover, PHA composed of 3HB-3HV monomers was synthesized in noanoic acid-grown PpUW2 (clone 19), as occurred in PpUW2 (clone 18). These data suggested that $PhaC_{19}$ was able to synthesize both SCL and SCL-MCL PHA.

Class I clone 25 synthesizes SCL and SCL-MCL PHA

Clone 25 contained 41,012-bp of DNA [GenBank: KT944269], which was predicted to originate from *β-Proteobacteria Variovorax* (Table 3). The *phaC*$_{25}$ and *phaA* genes were apparently located in a single operon (Fig. 3A). A *phaB* gene was located 9 open reading frames downstream of the *phaA*. The *phaC*$_{25}$*A* and *phaB* loci in clone 25 are similar to those in clone P1N3 (Supplementary 1), though the *phaB* gene in clone P1N3 was four open reading frames away from *phaC*$_{P1N3}$*A*. PhaC proteins of clones P1N3 and 25 shared 77% identity. In addition, $PhaC_{25}$ was 35% and 61% identical to $PhbC_{Sm}$ and $PhbC1_{Re}$. The conserved amino acid Asp130 affecting PhaC substrate specificity (Matsumoto et al. 2005) was replaced by Ala152 (Fig. 3B).

SCL-MCL PHA (C4-C6-C8) was produced in PpUW2 harbouring clone 25 when grown on octanoate, and only PHB (C4) was isolated from gluconic acid-grown cells (Table 4). Similar results were obtained in PpUW2 with clone 18 or clone 19 under the same conditions. However, the quantity of C6-C8 monomers was greater with clone 25 that those with clone 18 or 19. Nonanoic acid-grown PpUW2 (clone 25) produced 3HB-3HV copolymer similar to that from PpUW2 with clone 18 or clone 19 (Table 4). These data suggested that $PhaC_{25}$ had similar properties to those of $PhaC_{18}$ and $PhaC_{29}$, but different from that of $PhaC_{14}$.

Complementation of *S. meliloti phaC* mutant with 11AW PHA^+ clones

In an earlier study, *S. meliloti* Rm11476 (*exoY*::Tn5 *phaC*::Tn5-233) had been employed for functionally harvesting metagenomic library clones pCX92, pCX9M1, pCX9M3 and pCX9M5 expressing Class I PHA synthases (Schallmey et al. 2011). The PhaA/PhbB and PhaB/PhbB proteins encoded by the cloned metagenomic DNA and/or *S. meliloti* genome could supply (R)-3-hydroxybutyrate for PHB biosynthesis catalyzed by the PhaC/PhbC enzymes. In order to examine the possibility that the PHA^+ clones of 11AW metagenomic library were able to complement the PHA^- phenotype of *S. meliloti* Rm11476, The Class I clones 14, 18, 19 and 25 (Table 3) were introduced into *S. meliloti* Rm11476. In each case, complementation was observed, indicating PhaC function in the *S. meliloti* background (Fig. 4), with greatest accumulation of PHA observed with clone 14. However, when clones pCX92, pCX9M1, pCX9M3 and pCX9M5 were transferred to *P. putida* PpUW2 (PHA^-), PHA was not detected in the recombinant strains grown with either octanoate or gluconic acid (data not shown). These results strengthen the argument that multiple hosts should be employed to screen for novel *phaC* genes.

We then asked whether Class II *phaC* could function in the *S. meliloti* background. First, we observed no detection of PHA when $phaC1_{Pp}$ (P*lac*::$phaC1_{Pp}$) was expressed in Rm11476 even though the *lac* promoter is functional and 3-hydroxybutyrate is supplied by the $PhbAB_{Sm}$. Similarly, PHA^+ clones 1 and 20 carrying Class II *phaC* genes also failed to complement the PHB^- phenotype of *S. meliloti* Rm11476 (Fig. 4), implying that the $phaC2_1$ and $phaC_{20}$ genes were not expressed and/or the substrate (R)-3-hydroxyalkanoic acids were absent in *S. meliloti*. However, when clone 16 was introduced into *S. meliloti* Rm111476, PHA was accumulated to a similar level as with Rm11476 containing PHB^+ clones pCX92, pCX9M1, pCX9M3 or pCX9M5 (Fig. 4). These data indicated that the $phaC1_{16}$ gene was expressed and its product was functional. These results suggest that *P. putida* is a more permissive surrogate host than *S. meliloti* for screening of novel PHA synthases.

Discussion

Biosynthesis of SCL-MCL PHA requires PHA synthases having a broad range of substrate specificity and SCL and MCL precursors (R-3-hydroxyacyl-CoAs) must be available within the cell. Building on earlier work that has sought to mine metagenomic DNA for modification of bacterial PHA production, we have successfully isolated cosmid clones that are able to functionally complement a *P. putida* PHA synthesis mutant. DNA sequence analysis revealed that the isolated metagenomic DNA originated from a broad diversity of bacteria, and encoded either Class I or Class II PHA synthases enzymes. Of note is that DNA from most of the isolated clones did not very closely match known sequences. Also of interest is the influence that the clones had on the quality and quantity of PHA produced in the *P. putida* surrogate host background.

A major appeal of PHA is that, as alternatives to conventional fossil fuel-derived plastics, they can be produced from renewable resources. The potential to improve the efficiency and cost of production, and expand the range of polymers and copolymers that are available for production, should impact on competitiveness and adoption of these materials in the marketplace. In this context, it is important to understand how the properties of PHA are determined by the PHA synthase enzyme and the substrate that is available to those enzymes, which is in turn influenced by the culture conditions and the metabolic pathways leading to substrate formation (Meng et al. 2014). PHA consisting of both short-chain-length (SCL, ≤C5) and medium-chain-length (MCL, ≥C6) monomers have great flexibility, decreased breakage and reduced melting point (Noda et al. 2005). Most naturally occurring PhaC proteins polymerize either SCL- or MCL-monomers, but there are a few know examples such as the Class I PhaC of *Aeromonas caviae* (Fukui and Doi 1997) and some *Pseudomonas* Class II PhaC (Matsusaki et al. 1998) that can synthesize SCL-MCL copolymers.

In the present work, heterologous complementation of the *P. putida* PHA$^-$ strain facilitated efficient simultaneous screening of millions of 11AW metagenomic clones, resulting in recovery of a greater number of novel Class I, Class II, and unclassified PHA synthases than from previous reports (Schallmey et al. 2011; Cheema et al. 2012). Additionally, the clones carrying Class I *phaC* isolated in *P. putida* were able to complement the PHA$^-$ phenotype of *S. meliloti*, but the previously isolated Class I clones isolated in *S. meliloti* (Schallmey et al. 2011) failed to synthesize PHA in *P. putida.* These results support the development of multi-host systems to increase the chances of the successful expression of *pha* genes of interest.

Biosynthesis of SCL-MCL PHA of various forms by Class I clones 14, 18, 19 and 25 and Class II clones 16 and 20 suggested availability of a full range of R-3-hydroxyacyl-CoAs derived from β-oxidation of fatty acids such as nonanoic acid and octanoate. The differential monomer

509 compositions in PHA synthesized by the Class II enzymes $PhaC_{16}$ and $PhaC_{20}$ are of particular
510 interest, especially since most Class II PHA synthases are unable to incorporate SCL monomers such
511 as 3HB and 3HV. Changes of conserved amino acid residues Ser325Thr, Leu484Val, and
512 Glu508Leu (conserved positions in *P. putida* PhaC1) have been identified to be involved in PhaC
513 substrate specificity and PHA yield (Takase et al. 2003; Shozui et al. 2010; Chen et al. 2014).
514 Presence of the amino acids Thr353 (corresponding to *P. putida* PhaC1 position 325), Val516
515 (corresponding to *P. putida* PhaC1 position 484) and Ala540 (corresponding to *P. putida* PhaC1
516 position 508) $PhaC_{20}$ might contribute to the enzyme's ability to incorporate SCL monomers
517 compared to those of 11AW $PhaC2_1$ and PHA synthases of *P putida* KT2440. Similarly, the
518 presence of three residues (Thr317, Val478 and Leu502) at these positions in $PhaC1_{16}$, most likely
519 resulted in increased 3HB content as demonstrated previously with engineered Class II PHA
520 synthases (Takase et al. 2003; Shozui et al. 2010; Chen et al. 2014).
521
522 **Conclusions**
523 We obtained 27 clones encoding Class I, II and unclassified PHA synthases by functional
524 screening of soil metagenomic cosmid clones in a *P. putida* PHA^- strain. Seven clones that were
525 characterized in more detail were able to produce a broad range of polymers and copolymers,
526 including SCL-MCL mixtures, depending on carbon source. Through this work we have
527 demonstrated the potential for using metagenome-derived clones for production of a variety of PHAs
528 of possible industrial utility. The collection of PHA metabolism genes from uncultivated organisms
529 provides not only a resource for production strain development, but also a series of sequence
530 templates that could prove useful in enzyme engineering efforts directed towards generation of PHA
531 products with desired properties.

532

533 **Nucleotide sequence accession numbers**

534 Complete sequences of 11AW metagenomic PHA$^+$ DNA have been deposited in GenBank
535 nucleotide sequence database [GenBank: KT944254-KT944278, KU728995 and KU728996].

536

537 **Abbreviations**

538 *P. putida*, *Pseudomonas putida*; PHA, polyhydroxyalkanoate; PhaA/PhbA, β-ketothiolase;
539 PhaB/PhbB, acetoacetyl-CoA reductase; PhaC/PhbC, PHA synthase; PhaZ, PHA depolymerase;
540 PhaJ, R-specific enoyl-CoA hydratase; 3HA, 3-hydroxyalkanoate; 3HB, 3-hydroxybutyrate; 3HV.
541 3-hydroxyvalerate; 3HHx, 3-hydroxyhexanoate; 3HP, 3-hydroxyheptanoate; 3HO, 3-
542 hydroxyoctanoate; 3HN, 3-hydroxynonanoate; 3HD, 3-hydroxydecanoate; SCL, short chain length;
543 MCL, medium chain length; *S. meliloti*, *Sinorhizobium meliloti*; GC-MS, gas chromatography-mass
544 spectrometry; CDW, cell dry weight.

545

546 **Authors's contributions**

547 JC and TCC conceived and designed this study, carried out all the experiments, and prepared
548 the manuscript. Both authors read and approved the final manuscript.

549

550 **Acknowledgements**

551 We are very grateful to Ricardo Nordeste for assisting with GC-MS analysis of PHA. We
552 wish to thank Dr. Bruce Ramsay (Polyferm Canada) for providing PHA standards, and Dr. Josh
553 Neufeld for use of the FilterMax F5 Multi-mode microplate reader.

554

Funding

This work was financially supported by a New Directions grant from Ontario Ministry of Agriculture, Food, and Rural Affairs (award number 381646-09), by a Strategic Projects grant from the Natural Sciences and Engineering Research Council of Canada (award number ND2012-1679), and by Genome Canada, through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled "Microbial Genomics for Biofuels and Co-Products from Biorefining Processes" awarded to David Levin and Richard Sparling.

Conflict of interest

Trevor Charles received partial financial support for presenting some of this material at the Society for Industrial Microbiology and Biotechnology Annual Meeting in 2014 and 2015. Jiujun Cheng declares no conflict of interest.

Ethical statement

The authors certify that this manuscript has not been published previously, and not under consideration for publication elsewhere, in whole or in part. No data have been fabricated or manipulated (including images), and no data, text, or theories by others are presented as if they were the authors' own. Consent to submit has been received explicitly from all the authors listed. And authors whose names appear on the submission have contributed sufficiently to the scientific work and therefore share collective responsibility and accountability for the results. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Cheema S, Bassas-Galia M, Sarma PM, Lal B, Arias S (2012) Exploiting metagenomic diversity for novel polyhydroxyalkanoate synthases: Production of a terpolymer poly(3-hydroxybutyrate-co-3-hydroxyhexanoate-co-3-hydroxyoctanoate) with a recombinant *Pseudomonas putida* strain. Bioresource Technology 103:322–328. doi: 10.1016/j.biortech.2011.09.098

Chen Y-J, Tsai P-C, Hsu C-H, Lee C-Y (2014) Critical residues of class II PHA synthase for expanding the substrate specificity and enhancing the biosynthesis of polyhydroxyalkanoate. Enzyme and Microbial Technology 56:60–66. doi: 10.1016/j.enzmictec.2014.01.005

Cheng J, Pinnell L, Engel K, Neufeld JD, Charles TC (2014) Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. Journal of Microbiological Methods 99:27–34. doi: 10.1016/j.mimet.2014.01.015

de Eugenio LI, Escapa IF, Morales V, Dinjaski N, Galàn B, García JL, Prieto MA (2010) The turnover of medium-chain-length polyhydroxyalkanoates in *Pseudomonas putida* KT2442 and the fundamental role of PhaZ depolymerase for the metabolic balance. Environ Microbiol 12:207–221. doi: 10.1111/j.1462-2920.2009.02061.x

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113–19. doi: 10.1186/1471-2105-5-113

Escapa IF, Morales V, Martino VP, Pollet E, Avérous L, García JL, Prieto MA (2011) Disruption of β-oxidation pathway in *Pseudomonas putida* KT2442 to produce new functionalized PHAs with thioester groups. Appl Microbiol Biotechnol 89:1583–1598. doi: 10.1007/s00253-011-3099-4

Fellay R, Frey J, Krisch H (1987) Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vitro insertional mutagenesis of gram-negative bacteria. Gene 52:147–154.

Foong CP, Lau N-S, Deguchi S, Toyofuku T, Taylor TD, Sudesh K, Matsui M (2014) Whole genome amplification approach reveals novel polyhydroxyalkanoate synthases (PhaCs) from Japan trench and nankai trough seawater. BMC Microbiol 14:e11836. doi: 10.1186/s12866-014-0318-z

Fukui T, Doi Y (1997) Cloning and analysis of the poly(3-hydroxybutyrate-co-3-hydroxyhexanoate) biosynthesis genes of *Aeromonas caviae*. J Bacteriol 179:4821–4830. doi: 10.1186/s12866-014-0318-z

Fukui T, Shiomi N, Doi Y (1998) Expression and characterization of (R)-specific enoyl coenzyme A hydratase involved in polyhydroxyalkanoate biosynthesis by Aeromonas caviae. J Bacteriol 180:667–673.

Hoppensack A, Rehm BH, Steinbüchel A (1999) Analysis of 4-phosphopantetheinylation of polyhydroxybutyrate synthase from *Ralstonia eutropha*: generation of beta-alanine auxotrophic Tn5 mutants and cloning of the *panD* gene region. J Bacteriol 181:1429–1435.

617 Jia Y, Kappock TJ, Frick T, Sinskey AJ, Stubbe J (2000) Lipases provide a new mechanistic model
618 for polyhydroxybutyrate (PHB) synthases: characterization of the functional residues in
619 *Chromatium vinosum* PHB synthase. Biochemistry 39:3927–3936. doi: 10.1021/bi9928086

620 Kawashima Y, Cheng W, Mifune J, Orita I, Nakamura S, Fukui T (2012) Characterization and
621 functional analyses of R-specific enoyl coenzyme A hydratases in polyhydroxyalkanoate-
622 producing *Ralstonia eutropha*. Appl Environ Microbiol 78:493–502. doi: 10.1128/AEM.06937-
623 11

624 Keshavarz T, Roy I (2010) Polyhydroxyalkanoates: bioplastics with a green agenda. Current
625 Opinion in Microbiology 13:321–326. doi: 10.1016/j.mib.2010.02.006

626 Lam KN, Hall MW, Engel K, Vey G, Cheng J, Neufeld JD, Charles TC (2014) Evaluation of a
627 pooled strategy for high-throughput sequencing of cosmid clones from metagenomic libraries.
628 PLoS ONE 9:e98968. doi: 10.1371/journal.pone.0098968

629 Matsumoto K, Takase K, Aoki E, Doi Y, Taguchi S (2005) Synergistic effects of Glu130Asp
630 substitution in the type II polyhydroxyalkanoate (PHA) synthase: enhancement of PHA
631 production and alteration of polymer molecular weight. Biomacromolecules 6:99–104. doi:
632 10.1021/bm049650b

633 Matsusaki H, Manji S, Taguchi K, Kato M, Fukui T, Doi Y (1998) Cloning and molecular analysis
634 of the Poly(3-hydroxybutyrate) and Poly(3-hydroxybutyrate-co-3-hydroxyalkanoate)
635 biosynthesis genes in *Pseudomonas* sp. strain 61-3. J Bacteriol 180:6459–6467.

636 Meng D-C, Shen R, Yao H, Chen J-C, Wu Q, Chen G-Q (2014) Engineering the diversity of
637 polyesters. Curr Opin Biotechnol 29:24–33. doi: 10.1016/j.copbio.2014.02.013

638 Noda I, Green PR, Satkowski MM, Schechtman LA (2005) Preparation and properties of a novel
639 class of polyhydroxyalkanoate copolymers. Biomacromolecules 6:580–586. doi:
640 10.1021/bm049472m

641 Park SJ, Kim TW, Kim MK, Lee SY, Lim S-C (2012) Advanced bacterial polyhydroxyalkanoates:
642 towards a versatile and sustainable platform for unnatural tailor-made polyesters. Biotechnology
643 Advances 30:1196–1206. doi: 10.1016/j.biotechadv.2011.11.007

644 Patil KR, Roune L, McHardy AC (2012) The PhyloPythiaS web server for taxonomic assignment of
645 metagenome sequences. PLoS ONE 7:e38581. doi: 10.1371/journal.pone.0038581

646 Pärnänen K, Karkman A, Virta M, Eronen-Rasimus E, Kaartokallio H (2015) Discovery of bacterial
647 polyhydroxyalkanoate synthase (PhaC)-encoding genes from seasonal Baltic Sea ice and cold
648 estuarine waters. Extremophiles 19:197–206. doi: 10.1007/s00792-014-0699-9

649 Quandt J, Hynes MF (1993) Versatile suicide vectors which allow direct selection for gene
650 replacement in Gram-negative bacteria. Gene 127:15–21. doi: 10.1016/0378-1119(93)90611-6

651 Rehm BHA (2003) Polyester synthases: natural catalysts for plastics. Biochem J 376:15–33. doi:
652 10.1042/BJ20031254

653 Sato S, Kanazawa H, Tsuge T (2011) Expression and characterization of (R)-specific enoyl
654 coenzyme A hydratases making a channeling route to polyhydroxyalkanoate biosynthesis in
655 *Pseudomonas putida*. Appl Microbiol Biotechnol 90:951–959. doi: 10.1007/s00253-011-3150-5

656 Schallmey M, Ly A, Wang C, Meglei G, Voget S, Streit WR, Driscoll BT, Charles TC (2011)
657 Harvesting of novel polyhydroxyalkanaote (PHA) synthase encoding genes from a soil
658 metagenome library using phenotypic screening. FEMS Microbiol Lett 321:150–156. doi:
659 10.1111/j.1574-6968.2011.02324.x

660 Sheu DS, Wang YT, Lee CY (2000) Rapid detection of polyhydroxyalkanoate-accumulating
661 bacteria isolated from the environment by colony PCR. Microbiology 146 (Pt 8):2019–2025.
662 doi: 10.1099/00221287-146-8-2019

663 Shozui F, Sun J, Song Y, Yamada M, Sakai K, Matsumoto K, Takase K, Taguchi S (2010) A new
664 beneficial mutation in *Pseudomonas* sp. 61-3 polyhydroxyalkanoate (PHA) synthase for
665 enhanced cellular content of 3-hydroxybutyrate-based PHA explored using its enzyme homolog
666 as a mutation template. Biosci Biotechnol Biochem 74:1710–1712. doi: 10.1271/bbb.100224

667 Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. Appl Environ Microbiol
668 77:1153–1161. doi: 10.1128/AEM.02345-10

669 Steinbüchel A, Lütke-Eversloh T (2003) Metabolic engineering and pathway construction for
670 biotechnological production of relevant polyhydroxyalkanoates in microorganisms. Biochemical
671 Engineering Journal 16:81–96. doi: 10.1016/S1369-703X(03)00036-6

672 Tai YT, Foong CP, Najimudin N, Sudesh K (2015) Discovery of a new polyhydroxyalkanoate
673 synthase from limestone soil through metagenomic approach. Journal of Bioscience and
674 Bioengineering. doi: 10.1016/j.jbiosc.2015.08.008

675 Takase K, Taguchi S, Doi Y (2003) Enhanced synthesis of poly(3-hydroxybutyrate) in recombinant
676 *Escherichia coli* by means of error-prone PCR mutagenesis, saturation mutagenesis, and in vitro
677 recombination of the type II polyhydroxyalkanoate synthase gene. J Biochem 133:139–145.

678 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary
679 Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. doi: 10.1093/molbev/mst197

680 Tortajada M, da Silva LF, Prieto MA (2013) Second-generation functionalized medium-chain-length
681 polyhydroxyalkanoates: the gateway to high-value bioplastic applications. Int Microbiol 16:1–15.
682 doi: 10.2436/20.1501.01.175

683 Tripathi L, Wu L-P, Dechuan M, Chen J, Wu Q, Chen G-Q (2013) *Pseudomonas putida* KT2442 as
684 a platform for the biosynthesis of polyhydroxyalkanoates with adjustable monomer contents and
685 compositions. Bioresource Technology 142:225–231. doi: 10.1016/j.biortech.2013.05.027

686 Verlinden RAJ, Hill DJ, Kenward MA, Williams CD, Radecka I (2007) Bacterial synthesis of
687 biodegradable polyhydroxyalkanoates. J Appl Microbiol 102:1437–1449. doi: 10.1111/j.1365-
688 2672.2007.03335.x

Wu H-A, Sheu D-S, Lee C-Y (2003) Rapid differentiation between short-chain-length and medium-chain-length polyhydroxyalkanoate-accumulating bacteria with spectrofluorometry. Journal of Microbiological Methods 53:131–135. doi: 10.1016/S0167-7012(02)00232-4

Zhu W, Lomsadze A, Borodovsky M (2010) Ab initio gene identification in metagenomic sequences. Nucleic Acids Research 38:e132–e132. doi: 10.1093/nar/gkq275

Figure legends

Fig. 1. Activity-based screening of 11AW metagenomic library clones carrying PHA biosynthesis genes in *P. putida* PpUW2 (PHA^-). PpUW2 recombinant strains were grown in 0.1N M63 medium containing 0.5% Na octanoate and Nile red (0.5 µg/ml). Cosmid pJC8 and pJC67 (P*lac*::*phaC1*$_{Pp}$) were used as negative and positive controls respectively. 11AW clones 14 and 16 were streaked on the same selection plate to verify the PHA^+ phenotype initially screened on LB with 0.5% (w/v) octanoate.

Fig. 2. Phylogenetic analysis of the polyhydroalkanoate synthases (PhaC) of 11AW metagenomic library clones. Protein sequence alignments were performed using MUSCLE (Edgar 2004). Neighbour-jointing phylogenetic trees were generated with MEGA6 (Tamura et al. 2013). GenBank accession numbers of 11AW PhaC proteins: $PhaC2_1$ (ALV86274), $PhaC_3$ (ALV86289), $PhaC_4$ (ALV86299), $PhaC_5$ (ALV86308), $PhaC_6$ (ALV86351), $PhaC_7$ (partial, ALV86358), $PhaC_8$ (ALV86364), $PhaC_{10}$ (partial, ALV86387), $PhaC_{14}$ (ALV86397), $PhaC1_{16}$ (ALV86417), $PhaC2_{16}$ (ALV86419), $PhaC_{18}$ (ALV86462), $PhaC_{19}$ (ALV86476), $PhaC_{20}$ (ALV86493), $PhaC_{21}$ (ALV86517), $PhaC_{23}$ (ALV86529), $PhaC_{25}$ (ALV86574), $PhaC_{27}$ (ALV86602), $PhaC1_{50}$ (ALV86626), $PhaC2_{50}$ (ALV86626), $PhaC_{51}$ (ALV86651), $PhaC_{P1N3}$ (ALV86715), $PhaC_{P2N8}$ (ALV86750), $PhaC_{P2N10}$ (partial, ALV86755), $PhaC_{P4N10}$ (ALV86768), $PhaC_{P8N7}$ (ALV86771).

714

715 Fig. 3. The *pha* genes and PhaC proteins in 11AW metagenomic DNA library clones. (A) The PHA^+
716 clones are classified into Class I and II based on the PhaC protein sequences. (B) Conserved amino
717 acids required for the activity of PHA synthases are indicated in closed triangles. Residues affecting
718 substrate specificity were marked by open triangles. The positions of amino acid residues were
719 numbered based on the sequence of $PhaC1_{16}$ [GenBank: ALV86417]. GenBank accession numbers
720 of PhaC proteins: $PhaC_{14}$, ALV86397; $PhaC_{18}$, ALV86462; $PhaC_{19}$, ALV86476; $PhaC_{25}$,
721 ALV86574; $PhaC1_2$, ALV86274; $PhaC_{20}$, ALV86493; $PhaC1_{16}$, ALV86417; *C. necator* PhbC1,
722 AAA21975; *P. putida* PhaC1, Q88D25; and *P.* sp. 61-3 PhaC1, BAA36200.

723

724 Fig. 4. Complementation of *S. meliloti* Rm11476 (*phbC*) with metagenomic clones. The cosmid
725 DNA of PHA^+ clones encoding Class I and II PHA synthases was transferred to *S. meliloti* Rm11476
726 by conjugation. Recombinant strains were grown in YM medium. PHA production was estimated by
727 measuring the fluorescence of Nile red stained cells. *S. meliloti* wild type Rm1021 and previously
728 isolated Class I clones (CX92, pCX92M1, pCX92M3, and pCX92M5; (Schallmey et al. 2011)) were
729 used as controls.

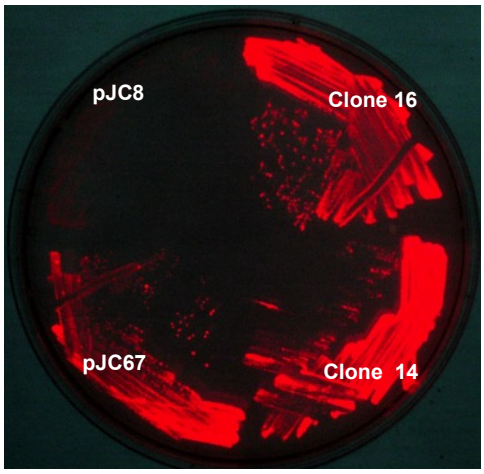


Figure 1.

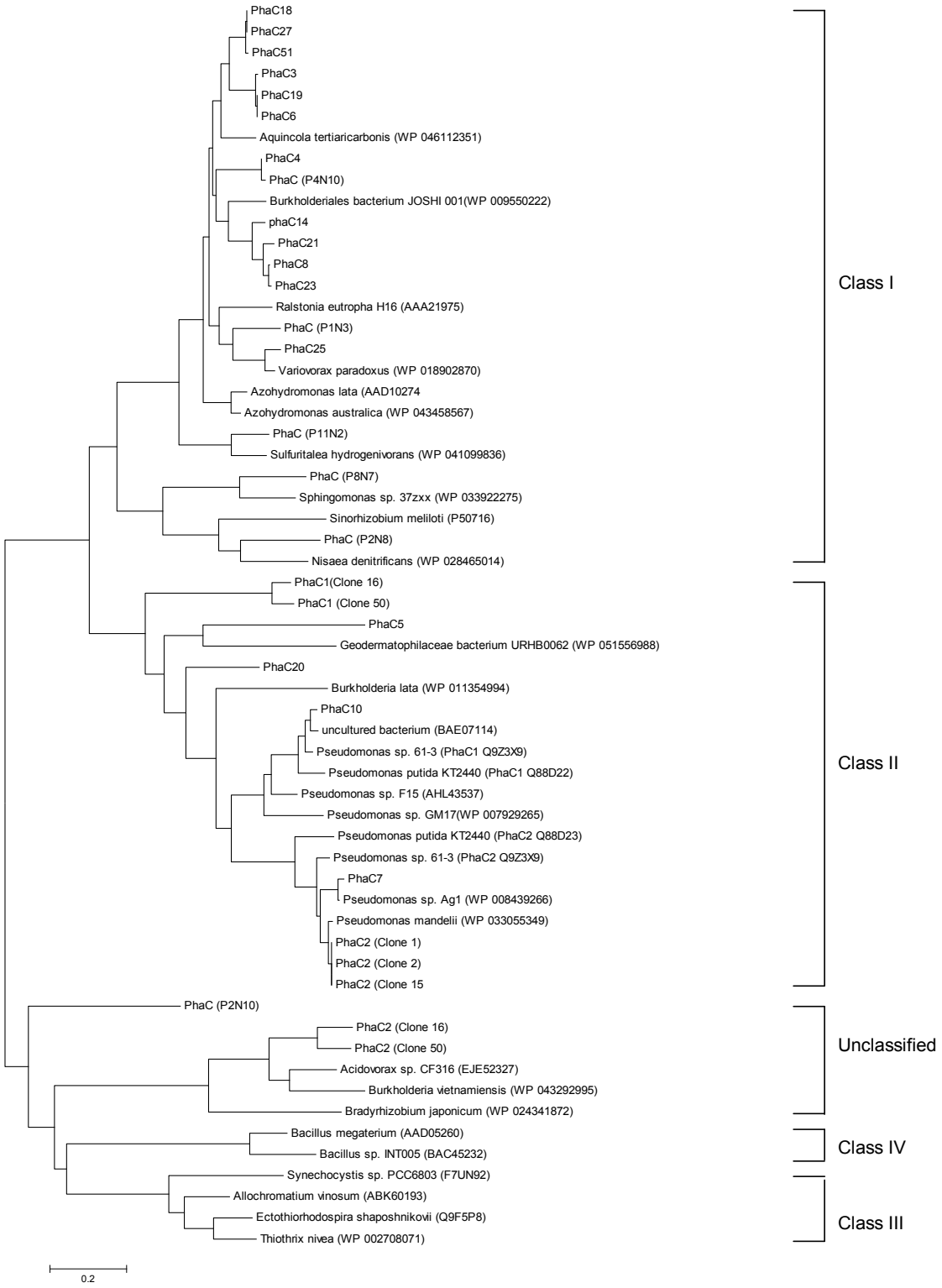
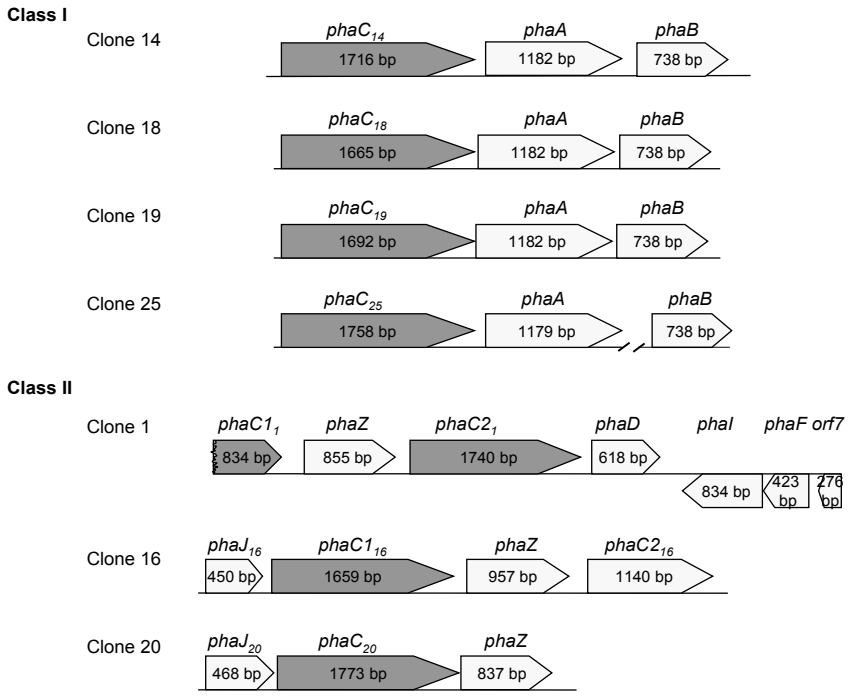


Fig. 2

A



B

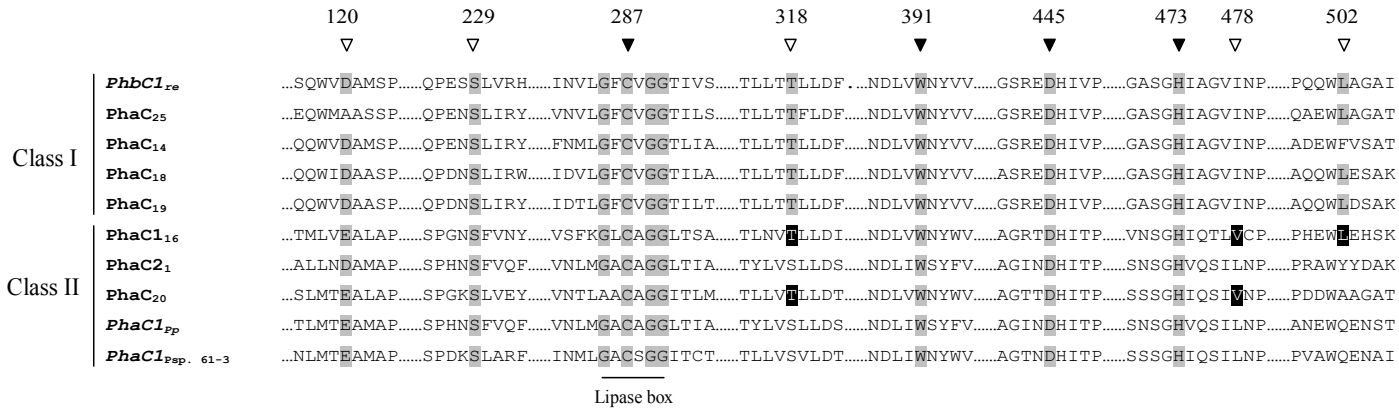


Fig. 3

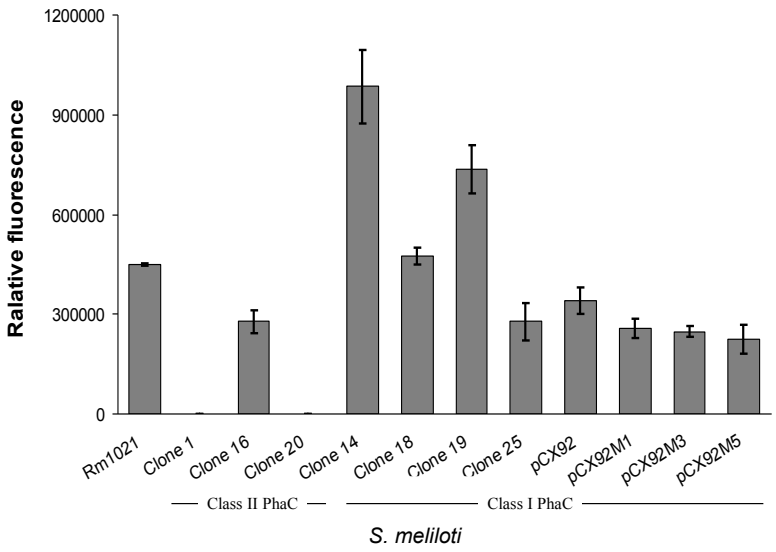


Fig. 4

Table 1. Bacterial strains, plasmids and cosmids

Strain, plasmids and cosmids	Relevant characteristics*	Reference
E. coli		
DH5α	F⁻ ϕ80 *lacZΔM15*Δ(*lacZYA-argF*)*U169 recA1 endA1 hsdR17 phoA supE44 thi-1 gyrA96 relA1*	Life Technologies
DH5α (Rif^R)	A spontaneous Rif^r mutant of DH5α	(Cheng et al. 2014)
HB101	F⁻ *thi-1 hsdS20 supE44 recA13 ara-14 leuB6 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1*, Sm^R	Life Technologies
S. meliloti		
Rm1021	Wild type, Sm^R	Lab collection
Rm11476	Rm1021, *exoY* ::Tn5 *phbC*::Tn5-233, Sm^R Nm^R	(Schallmey et al. 2011)
P. putida		
KT2440	Wild type	Lab collection
PpUW1	A spontaneous Rif^R mutant of KT2440, Rif^R	This study
PpUW2	PpUW1 (*phaC1-phaZ-phaC2*::ΩKm^R), Rif^R Km^R	This study
Plasmids/cosmids		
pRK600	pRK2013 *tra* Nm^R::Tn9, Cm^R	Lab collection
pRK7813	IncP *oriT cos lacZα*, Tc^R	Lab collection
pJC8	Broad-host-range Gateway® entry cosmid, Tc^R	(Cheng et al. 2014)
pJQ200-SK	Cloning vector, *sacB*, Gm^R	(Quandt and Hynes 1993)
pHP45-ΩKm^R	Ω-Km^R cassette, Km^R	(Fellay et al. 1987)
pJC63	pJQ200-SK carrying 5′-*phaC1* of *P. putida* KT2440, Gm^R	This study
pJC69	pJC63 carrying 3′-*phaC2* of *P. putida* KT2440, Gm^R	This study
pJC71	pJC69 carrying Ω-Km^R between the *phaC1′* and *phaC2′* fragments, Gm^R Km^R	This study
pJC67	pRK7813 carrying P*lac*::*phaC1*$_{Pp}$, Tc^R	This study
pJC123	pRK7813 carrying P*lac*::*phaC1*$_{16}$, Tc^R	This study
pJC157	pRK7813 carrying P*lac*::*phaC1*$_{16}$::*phaC2*$_{16}$, Tc^R	This study
pJC159	pRK7813 carying P*lac*::*phaC2*$_{16}$, Tc^R	This study
pCX92	pRK7813 carrying *phaC* and *phaP*, Tc^R	(Schallmey et al. 2011)
pCX9M1	pRK7813 carrying *phaC*, *phaA* and *phaB,* Tc^R	(Schallmey et al. 2011)
pCX9M3	pRK7813 carrying *phaC*, Tc^R	(Schallmey et al. 2011)
pCX9M5	pRK7813 carrying *phaC* and *phaB*, Tc^R	(Schallmey et al. 2011)

*Antibiotic resistances: Cm^r, chloramphenicol; Gm^R, gentamycin; Km^R, kanamycin; Nm^R, neomycin; Sm^R, streptomycin; Tc^R, tetracycline.

Table 2. DNA oligonucleotides

Oligo ID	Sequence
JC159	CGCCAAGCTTGT**TGA**GGAGCGTCGTAGATGAGTAAC
JC160	GCGCGGATCCTCAACGCTCGTGAACGTAGGT
JC161	GCGCTCTAGACACGAAAGTCGGGTAACCAAT
JC162	GGTCGGATCCCGCCAGCTGATGATGAAGGTC
JC163	GCTGGGATCCGTTGTGGCTGGAGTGGATCAC
JC164	CGCCCTCGAGCACCGGTACCCTGATCTGATA
phaCF1	ATCAACAARTWCTACRTCYTSGACCT
phaCR2	GTSTTCRTSRTSWSCTGGCGCAACCC
phaCR4	AGGTAGTTGTYGACSMMRTAGKTCCA
JC202	CCGGATGCA**TGA**TAGGACGCCGGATGACAGCAGAGAAGGCTGAG
KC228	CCCAGATCTGCAGTATTCTAGATTCAGGCGCGCACGCGCACGTA
JC204	CCGGTCTAGAGGAACGTCCCATGAGCAATCC
JC205	GGGGATGCATGCCTCAGGCGAAGTACTGGCCGCCATTG
JC301	CGCGAAGCTTC**TGA**AGGAGAGAACACATGGCGAAGACCA
JC302	GGGGAATTCTGAATATCAGCGTCGGAGCATG
JC333	CGCGAAGCTTGT**TGA**GGAGAGATCTATGAGCTCCCTCCCCGTGGCCGAACGTTC
JC334	CGCTGAATTCTCTACGCCGCCTGACCCACGAT
JC337	ACCAGAATTCAGGAGATCTATATGAGCTCCCTCCCCGTGGCCGAAC
pJET forward	CGACTCACTATAGGGAGAGCGGC
pJET reverse	AAGAACATCGATTTTCCATGGCAG

Restriction sites are underlined. Translation stop codon is in bold, and ribosomal binding site is double underlined.

Table 3. 11AW metagenomic clones encoding PHA synthases (PhaC).

PHA$^+$ Clones	DNA (bp)	GC content (%)	Possible origin of cloned DNA	Best match of PHA synthases	Identity (%)	Cosmid Genbank ID
Class I						
3	>16,077	67.5	*Leptothrix* (*á-Proteobacteria*)	Hypothetical protein (WP_046112351), *Aquincola tertiaricarbonis*	74	KT944255
4	40,585	65.7	*Leptothrix* (*á-Proteobacteria*)	Polyhydroxyalkanoate synthase (AAD10274), *Azohydromonas lata*	72	KT944256
6	>11,479	67.7	*Leptothrix* (*á-Proteobacteria*)	Polyhydroxyalkanoate synthase (AAD10274), *Azohydromonas lata*	73	KT944258
8	34,220	67.8	*Leptothrix* (*á-Proteobacteria*)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	74	KT944260
14	29,565	69.1	*Methylibium* (*á-Proteobacteria*)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	74	KT944262
18	37,718	69.4	*Rubrivirax* (*â-Proteobacteria*)	Hypothetical protein (WP_043458567), *Azohydromonas austrilica*	73	KT944264
19	25,085	68.1	*Leptothrix* (*á-Proteobacteria*)	Polyhydroxyalkanoate synthase (AAD10274), *Azohydromonas lata*	73	KT944265
21	33,064	69.3	*Gemmatimonas* (*Gemmatimonadetes*)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	73	KT944267
23	42,865	69.6	*Leptothrix* (*á-Proteobacteria*)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	74	KT944268
25	41,012	66.8	*Varivorax* (*á-Proteobacteria*)	Poly-(R)-hydroxyalkanoic acid synthase (WP_018902870), *Variovorax paradoxus*	93	KT944269
27	37,156	67.0	*Rubrivivax* (*á-Proteobacteria*)	Hypothetical protein (WP_043458567), *Azohydromonas austrilica*	73	KT944270
51	38,794	69.9	*Rubrivirax* (*á-Proteobacteria*)	Hypothetical protein (WP_043458567), *Azohydromonas australica*	73	KT944272
P2N8	22,625	66.3	*Xanthomondaceae* (*ã-Proteobacteria*)	Poly(3-hydroxyalkanonate) synthase (WP_028465014), *Nisaea denitrificans*	55	KT944272
P11N2	40,293	58.4	*Janthinobacterium* (*á-Proteobacteria*)	Poly(3-hydroxyalkanonate) synthase (WP_041099836), *Sulfuritalea hydrogenivorans*	73	KT944278
P1N3	32,797	68.6	*Variovorax* (*á-Proteobacteria*)	Poly(3-hydroxyalkanonate) synthase (WP_028253507), *Variovorax* sp. URB0020	86	KT944273
P4N10	>2,724	64.2	*Leptothrix* (*á-Proteobacteria*)	Polyhydroxyalkanonate synthase (AAD10274), *Azohydromonas lata*	72	KT944276
P8N7	>5,619	67.6	*Sphingomonadaceae* (*á-Proteobacteria*)	Poly-(R)-hydroalkanoic acid synthase (WP_033922275), *Sphingomonas* sp. 37zxx	67	KT944277
ClassII						
16 (PhaC1$_{16}$)	33,900	66.1	*Methylibium* (*á-Proteobacteria*)	PhaC1 synthase (AHL43537), *Pseudomonas* sp.F15	48	KT944263
50 (PhaC1$_{50}$)	22,965	67.0	*Hydrocarboniphaga* (*á-Proteobacteria*)	Polyhydroxyalkanoic acid synthase (WP_007929265), *Pseudomonas* sp. GM17	46	KT944271
1 (PhaC1$_{2}$)	42,747	60.0	*Pseudomonas* (*ã-Proteobacteria*)	Poly(R)-hydroxyalkanoic acid synthase (WP_033055349). *Pseudomonas mandelii*	98	KT944254
2	59,765	59.1	*Pseudomonas* (*ã-Proteobacteria*)	Poly(R)-hydroxyalkanoic acid synthase (WP_033055349). *Pseudomonas mandelii*	98	KU729001
15	59,765	59.1	*Pseudomonas* (*ã-Proteobacteria*)	Poly(R)-hydroxyalkanoic acid synthase (WP_033055349). *Pseudomonas mandelii*	98	KU728996
5	37473	66.3	*Burkholderia* (*á-Proteobacteria*)	Poly-beta-hydroxybutyrate polymerase (WP_051556988), *Geodermatophilaceae bacterium* URHB0062	53	KT944257
7	>407	65.6	*Pseudomonadaceae* (*ã-Proteobacteria*)	Poly(R)-hydroxyalkanoic acid synthase (WP_008439266), *Pseudomonas* sp. Ag1	100	KT944259
10	>404	63.9	*Pseudomonas* (*ã-Proteobacteria*)	PHA synthase (BAE07114), uncultured bacterium	98	KT944261
20	30,964	66.8	*Salinibacter* (*Bacteroidetes*)	Poly(R)-hydroxyalkanoic acid synthase (WP_011354994). *Burkholderia lata*	59	KT944266
Unclassified						
P2N10	>13,508	66.6	*Alcagenaceae* (*á-Proteobacteria*)	Poly-R-hydroxyalkonic acid synthease PhaE (EGV16635), *Thiocapsa marina* 5811	27	KT944275
				Poly-beta-hydroxybutyrate polymerase PhaC (WP_002708071), *Thiothrix nivea*	43	KT944275
16 (PhaC2$_{16}$)	33,900	66.1	*Methylibium* (*á-Proteobacteria*)	Easterase (WP_043292995), *Burkholderia vietnamiensis*	53	KT944263
50 (PhaC2$_{50}$)	22,965	67.0	*Hydrocarboniphaga* (*ã-Proteobacteria*)	Easterase (WP_043292995), *Burkholderia vietnamiensis*	53	KT944271

The PhaC proteins were classified based on their amino acid sequences. Possible origin of cloned metagenomic DNA was predicted by PhyloPythia.

Table 4. Biosynthesis of SCL- and/or -MCL PHAs by 11AW metagenomic clones in *Pseudomonas putida*.

P. putida	Carbon sources	PHA content (%)	PHA composition (mol%)						
			3HB	3HV	3HHx	3HP	3HO	3HN	3HD
Class II									
PpUW1	Octanoate	30.0			0.7		99.3		
	Nonanoic acid	45.0				19.1		80.9	
	Gluconic acid	1.7					21.1		76.8
PpUW2 (clone 1)	Octanoate	32.7			4.5		95.5		
	Nonanoic acid	23.4				42.0		58.1	
	Gluconic acid	8.2					20.4		79.6
PpUW2 (clone 16)	Octanoate	25.1	17.3		39.3		43.2		
	Nonanoic acid	46.5	3.4	40.9		55.4			
	Gluconic acid	4.5			47.6		52.4		
PpUW2 (pJC123)	Octanoate	17.5	13.9		30.3		51.3		
PpUW2 (pJC157)	Octanoate	32.5	14.5		33.9		49.3		
PpUW2 (clone 20)	Octanoate	17.6			17.3		82.7		
	Nonanoic acid	45.0		7.5		50.2		42.3	
	Gluconic acid	3.9			9.9		37.4		52.7
Class I									
PpUW2 (clone 14)	Octanoate	41.8	98.9		1.1				
	Nonanoic acid	59.4	6.8	93.2					
	Gluconic acid	81.2	97.9		0.8		1.3		
PpUW2 (clone 18)	Octanoate	6.4	97.9		0.5		2.4		
	Nonanoic acid	17.3	32.8	67.2					
	Gluconic acid	30.7	100.0						
PpUW2 (clone 19)	Octanoate	30.9	98.8		0.5		0.7		
	Nonanoic acid	38.1	20.9	79.2					
	Gluconic acid	12.3	100.0						
PpUW2 (PhaC25)	Octanoate	15.6	95.0		2.9		2.1		
	Nonanoic acid	43.1	25.2	74.9					
	Gluconic acid	4.1	100.0						

P. putida PpUW1 (PHA^+) and PpUW2 (PHA^-) carrying PHA^+ cosmids of 11AW metagenomic clones were grown at 30°C for 48 h in 0.1N M63 media supplemented with 0.5% octanoate, 0.5% nonanoic acid or 1% gluconic acid. PHA content was represented as relative to cell dry weight. Values were the averages of triplicate cultures. The standard deviations were <10% of the values. Abbreviations: 3HB, 3-hydroxybutyrate (C4); 3HHx, 3-hydroxyhexanoate (C6); 3HO, 3-hydroxyoctanoate (C8); 3HD, 3-hydroxydecanoate (C10).

Supplementary figures

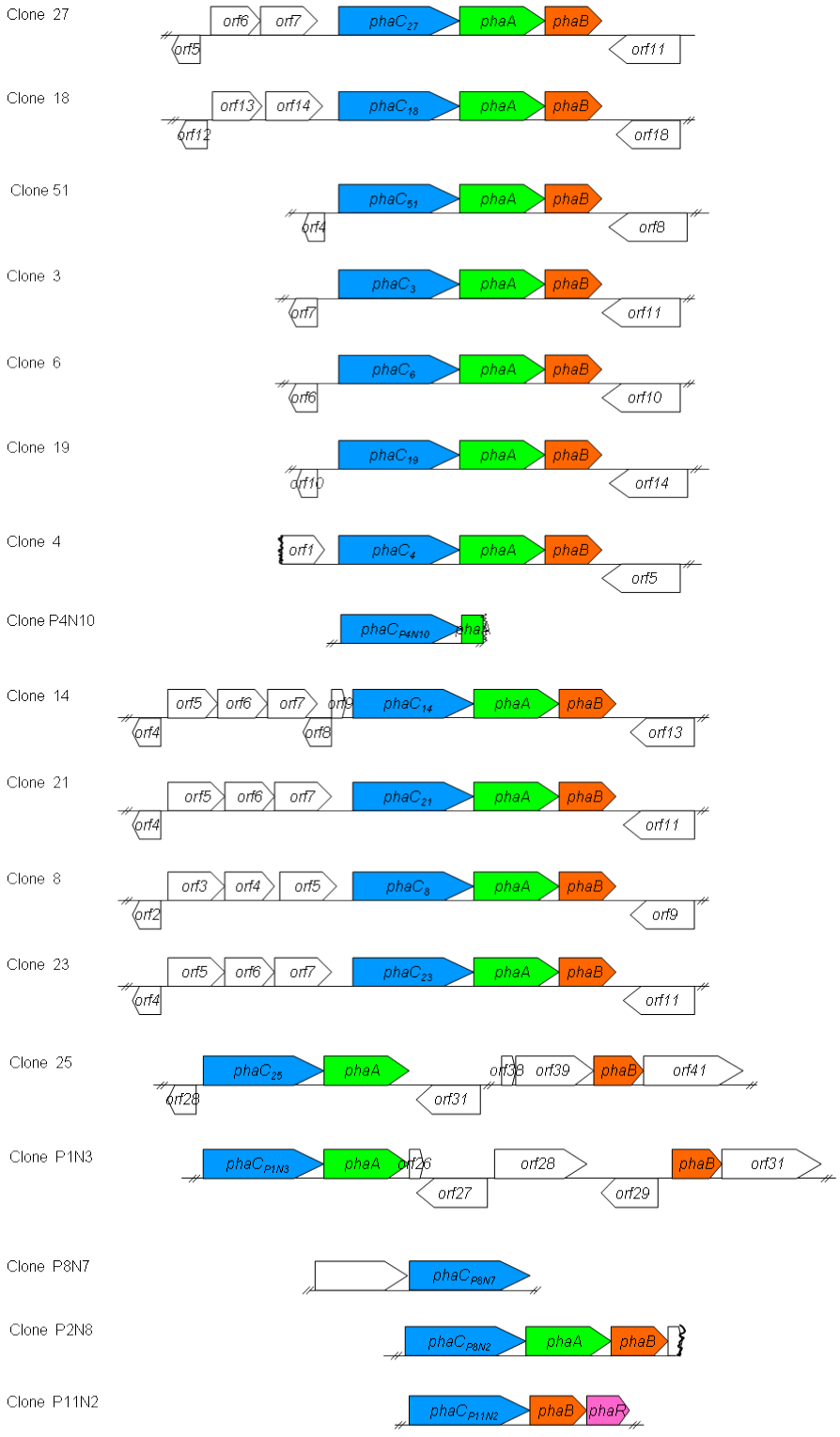
Supplementary 1. The *pha* gene locus in 11AW metagenomic DNA clones encoding PHA synthases (PhaC) were classified based on the PhaC protein sequences.

Supplementary 2. Transposon Tn4652 in 11AW PHA+ clones 2 and 15. Insertion of the Tn4652 into the *orf17* of clone 1 resulted in clone 2. Clone 15 was derived from clone 1 by the transposon insertion into the *orf19*.

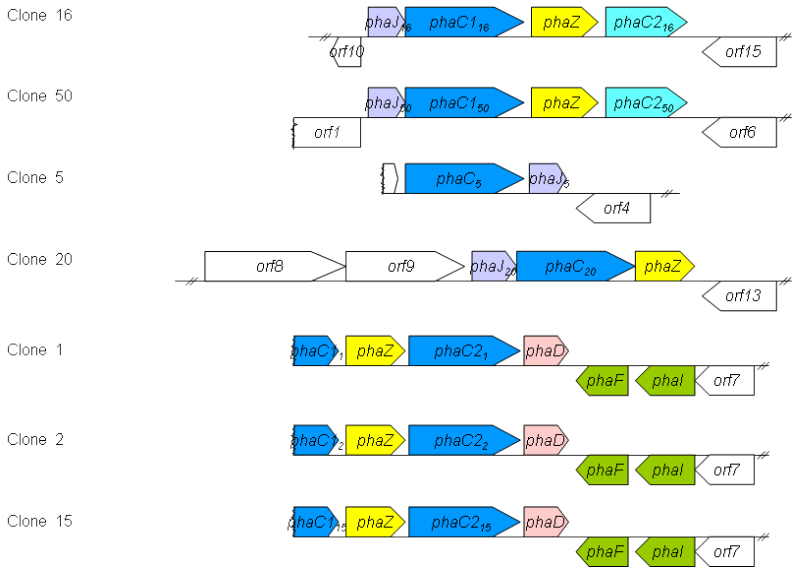
Supplementary 3. Multiple sequence alignment of PHA synthases of 11AW metagenomic DNA clones. Alignment was performed with MUSCLE. The conserved amino acid residues affecting substrate specificity in Class I and II PhaC are highlighted in purple and yellow respectively. The positions of amino acids essential for PhaC activity are highlighted in green. The position of amino acid residues are marked based on PhaC1$_{16}$ sequence. Names of Class I PhaC are in regular font, Class II PhaC in bold, and unclassified PhaC underlined. PhaC proteins with partial sequences are marked with a star.

Supplementary 4. R-specific enoyl-CoA hydratases PhaJ. (A) Phylogenetic tree was constructed with MEGA6. The bar represents substitution of amino acid residue. (B) Conserved amino acid Asp^{39}, His^{44} and Ser^{7} required for activity are marked.

Class I

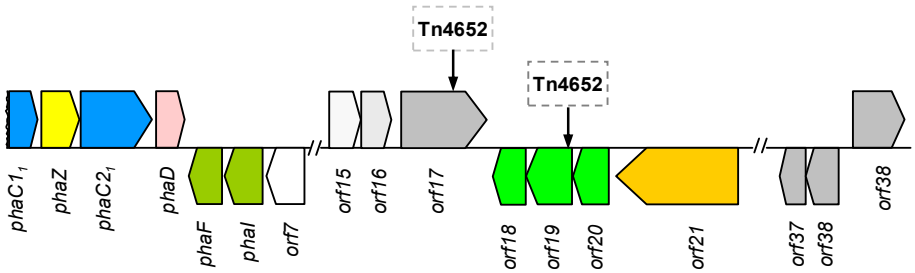


Class II



Unclassifed

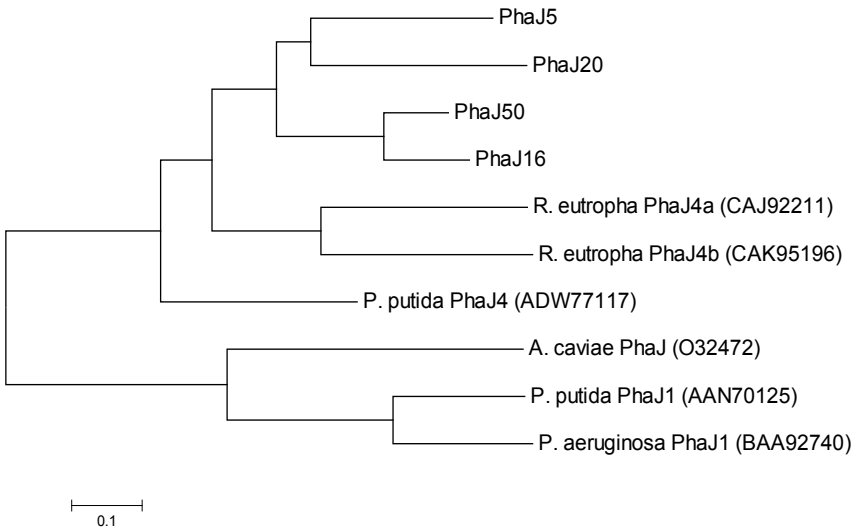
Supplementary 1.



Supplementary 2.

Supplementary 3

A.



B.

```
                                                                  39    44
R. eutropha PhaJ4a          -MrtiaSLeeLEglqGQEvavSDWIeVtQqqvnqFAdaTGDhQWIHIDVERAkKESPyGG
R. eutropha PhaJ4b          -MktyeNiadLQplVGevIGtSEWLaLDQARIntFAdaTGDhQWIHVDVERAkn-gPFGa
P. putida PhaJ4             --mphvpvteLsqyVGkELGhSEWLKIDQqRInlFAeaTGDfQfIHVDpEKAAK-tPFGG
PhaJ20                      mteEaySvrtIpqfaGrELGvSDWLRIDQeRIdrFAACTdDrQWIHVDVERAeRESPFGt
PhaJ50                      --mstlkMatLtekVGQELGttrWvaLDQSRIQaFAdCTGDqQWIHtDVERAkKESPFGG
PhaJ16                      --msaltMqnIaakIGeELGvtgWttLDQARIQaFAdCTGDgQWIHtDVERArKESPFGG
PhaJ5                       -MadaltLatIEsfVGrELGeSqWvRmDQqRIdaFAeCTGDrQWIHVDVERAAREgPFGG
A. caviae PhaJ              -------MsaqsleVGQkarlSk--RfgaAevaaFAAlseDfnplHlDpafAAt-taFer
P. putida PhaJ1             -MsQvtNtpyealeVGQkaeykk--sVEerdIQlFAAmsGDhnpvHlDaEfAAK-SmFre
P. aeruginosa PhaJ1         -MsQvqNipyaEleVGQkaeyts--sIaerdlQlFAAvsGDrnpvHlDaayAAt-tqFke

                                      71
R. eutropha PhaJ4a          PIAHGfLTLSLLPkfmhNalhmPskI--GVNYGLnRVRFtAPVpVGSklRaRIKLLkver
R. eutropha PhaJ4b          PIAHGfLTLSLLPafthsayrirns-stGVNYGLDKVRFpAPVpVdSllRaqfKLMSyea
P. putida PhaJ4             tIAHGfLTLSLiPklieDIlvlPqgLkMvVNYGLDsVRFiqPVKVdSrVRlkVKLgevve
PhaJ20                      tvAHGlLvLSLLPrfgfEVglvPpgVssalNYGfDgVRFvSPVKaGArVRdRVtLLeatd
PhaJ50                      PvAHGMLTLSLLPmwlfDLpaaPddagailNYGfDKVRFlAPVKsGAkVRaRIKLLAatP
PhaJ16                      PvAHGMLTLSLLPpwlfELpaaPldagailNYGfDKVRFlAPVKaGSrVRgRIKLLAaaP
PhaJ5                       tIAHGfLTLSiigpaqlDVwiaPagIgtaVhYGLDKVRFlAPVpaGrnVRtRIKLaAvea
A. caviae PhaJ              PIvHGML-LSLfsgllgqqlpgkgsIyLG-----qslsFklPVfVGdeVtaeVevtAlre
P. putida PhaJ1             rIAHGMfsgaLisaavactlpgPgtIyLG-----qQmsFqkPVKIGdtltvRleiLeklP
P. aeruginosa PhaJ1         rIAHGMLsgaLisaaiatVlpgPgtIyLG-----qtlRFtrPVKLGddlkvelevLeklP

R. eutropha PhaJ4a          ldplpkspeLvgaqsTwevTVErEGsdrPvcVAEsItrryg--------
R. eutropha PhaJ4b          len-------GgaqFkvemmVErQGgSKPvcIAEsILrrfP--------
P. putida PhaJ4             Kkp-------GqwLlkaiaTlEIEGEeKPAyIAEsLsLcfv--------
PhaJ20                      KgE-------GrlLvkarhTIEIEGESKPALVAEmLaMlItg--------
PhaJ50                      KeK-------GrvLlTqeyTVEIEnEtKPALIAElLVMlIPkaa-----
PhaJ16                      KdK-------GrlLlTqeyTVEIEnEtKPALIAElLVMlvPka------
PhaJ5                       Kgs-------GrtLvTtenvVEIEGhdKPALIAtaLaMiMP--------
A. caviae PhaJ              dkp---iatLttrIFTqggalaVtGE----------VVkLP--------
P. putida PhaJ1             KfK----vRiatnVYnqndelvVaGEAeilaprkqqtVeLvspPnfvas
P. aeruginosa PhaJ1         KnR----vRmatrVFnqagkqvVDGEAeimapeEkLsVeLaelPpisig
```

Supplementary 4