1   Functional metagenomics using *Pseudomonas putida* expands the known diversity of

2   polyhydroxyalkanoate synthases and enables the production of novel polyhydroxyalkanoate

3   copolymers

4

5   Jiujun Cheng and Trevor C. Charles*

6

7   Department of Biology and Centre for Bioengineering and Biotechnology, University of Waterloo,

8   Waterloo, Ontario, Canada N2L 3G1

9

10   *Corresponding author:

11   Dr. Trevor C. Charles

12   Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

13   E-mail: tcharles@uwaterloo.ca

14

15   Email addresses: Jiujun Cheng (j37cheng@uwaterloo.ca), Trevor C. Charles

16   (trevor.charles@uwaterloo.ca)

1

Highlights

17

18 • A functional metagenomic screen was developed for isolation of PHA synthesis genes using

19 *Pseudomonas putida*

20 • Use of the screen resulted in a collection of novel PHA synthesis genes that directed the

21 production of a variety of different polyhydroxyalkanaotes, including SCL-MCL copolymers

22 • DNA sequence analysis of the cloned DNA suggests some basis for the polyhydroxyalkanoate

23 diversity based on key amino acid residues in PHA synthase enzymes

24 • We have demonstrated the potential for using metagenome-derived clones for production of a

25 variety of PHAs of possible industrial utility.

26 **Abstract**

27 Bacterially produced biodegradable polyhydroxyalkanoates with versatile properties can be achieved

28 using different PHA synthase enzymes. This work aims to expand the diversity of known PHA

29 synthases via functional metagenomics, and demonstrates the use of these novel enzymes in PHA

30 production. Complementation of a PHA synthesis deficient *Pseudomonas putida* strain with a soil

31 metagenomic cosmid library retrieved 27 clones expressing either Class I, Class II or unclassified

32 PHA synthases, and many did not have close sequence matches to known PHA synthases. The

33 composition of PHA produced by these clones was dependent on both the supplied growth substrates

34 and the nature of the PHA synthase, with various combinations of SCL- and MCL-PHA. These data

35 demonstrate the ability to isolate diverse genes for PHA synthesis by functional metagenomics, and

36 their use for the production of a variety of PHA polymer and copolymer mixtures.

37 **Keywords**:

38 Functional metagenomics; soil metagenomic library; polyhydroxyalkanoate (PHA); bioplastics;

39 polyhydroxyalkanoate synthase; *Pseudomonas putida*; *Sinorhizobium meliloti*

40

2

1. 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 (Cheema et al., 2012; Keshavarz and Roy, 2010; Meng et al., 2014; Park et al., 2012; Schallmey et al., 2011; Tripathi et al., 2013).

PHA molecules consist of over 150 possible constituent monomers of hydroxyl fatty acids (Meng et al., 2014). 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.

63   The key Class I PHA synthesis pathway genes include *phaC* (or *phbC*), *phaA* (or *phbA*) and

64   *phaB* (or *phbB*). The three genes are often, but not always, clustered in a single operon. The

65   acetoacetyl-CoA reductase (EC 2.3.1.9) encoded by the *phaA* and β-ketothiolase (EC 1.1.1.36)

66   encoded by the *phaB* convert acetyl-CoA to (R)-3-hydroxybutyryl-CoA (3HB-CoA) through

67   acetoacetyl-CoA. The Class-I PhaC (EC 2.3.1.-) then polymerizes 3HB-CoA into

68   polyhydroxybutyrate (PHB) (Rehm, 2003). The Class II *pha* cluster is well conserved in

69   *Pseudomonas* and consists of two PHA synthase genes (*phaC1* and *phaC2*) flanking a PHA

70   depolymerase gene (*phaZ*), and *phaD* encoding a transcriptional activator of *pha* genes. Phasin-

71   encoding *phaF* and *phaI* gene are transcribed divergently to other *pha* genes. MCL monomers ((R)-

72   3-hydroxylacyl-CoA) are derived from β-oxidation of fatty acids or fatty acid *de novo* synthesis from

73   unrelated carbon sources (Tortajada et al., 2013).

74   Traditional culture-based strategies for obtaining new biocatalysts are limited by the inability

75   to cultivate the majority of environmental microbes. While sequence-based metagenomics can

76   identify genes homologous to those present in available sequence databases, it is difficult to reliably

77   predict the function of truly new genes through homology-based analysis. In contrast, functional

78   metagenomics involves the construction of gene libraries from microbial community genomic DNA

79   and functional screening for novel enzymes of interest or potential industrial applications (Simon

80   and Daniel, 2011). Functional metagenomics has the potential to identify truly novel sequences for a

81   given function.

82   In previous work, functional metagenomics was used to isolate new Class I PhaC from soil

83   metagenomic clones by Nile red staining and phenotypic screening in *α-Proteobacteria*

84   *Sinorhizobium meliloti* (Schallmey et al., 2011). In another study, *phaC* genes encoding both Class I

85   and II PhaC proteins were PCR amplified from oil-contaminated soil library clones (Cheema et al.,

86   2012). One of the isolated genes produced PHA copolymer when expressed in *Pseudomonas putida*.

87   Moreover, partial *phaC* genes were also obtained from metagenomic DNA via direct PCR

88   amplification (Foong et al., 2014; Pärnänen et al., 2015; Tai et al., 2015). Building on these previous

89   studies, we constructed a PHA⁻ strain of *P. putida* KT2440 to use as a surrogate host and

90   functionally identified *phaC* genes after screening millions of agricultural wheat soil metagenomic

91   clones. A total of 27 PHA⁺ clones were obtained. Accumulation and monomer composition of PHA

92   directed by seven of these clones were further examined.

93

94   **2. Methods**

95   **2.1. Bacterial strains, plasmids, cosmids and growth conditions**

96   Bacterial strains, plasmids and cosmids are listed in Table 1. *E. coli* strains were grown at

97   37°C in LB (Lennox) medium (1% tryptone (w/v), 0.5% yeast extract (w/v), and 0.5% NaCl (w/v),

98   pH 7). *Pseudomonas* strains were grown at 30°C in LB or 0.1N M63 minimal medium (Escapa et al.,

99   2011) supplemented with 0.5% sodium octanoate (w/v), 0.5% nonanoic acid (v/v) or 1% gluconic

100   acid (w/v). *S. meliloti* was grown in LB or YM medium (Schallmey et al., 2011). Antibiotics were

101   used at the following concentrations: streptomycin, 200 µg/ml for *S. meliloti* and 100 µg/ml for *E.*

102   *coli*; kanamycin, 100 µg/ml for *Pseudomonas* and 50 µg/ml for *E. coli*; neomycin, 200 µg/ml;

103   rifampicin, 100 µg/ml; gentamicin, 10 µg/ml for *E. coli* and 100 for *P. putida*; and tetracycline 20

104   µg/ml for *E. coli* or 40 µg/ml for *P. putida*.

105

106   **2.2. Construction of PHA⁻ strain PpUW2**

107   DNA oligonucleotides are listed in Table 2. A Rif^R spontaneous mutant PpUW1 of *P. putida*

108   KT2440 was generated by plating a culture of strain KT2440 on a LB Rif plate, followed by single

109   colony purification. To construct the PHA deficient strain PpUW2 (*phaC1ZC2*) of *P. putida* PpUW1,

110   a 943-bp DNA fragment containing 5´-*phaC1* region (766 bp) was PCR amplified using *P. putida*

111   KT2440 genomic DNA as a template and primer pair JC161-JC162, digested with HindIII-BamHI

112   and then cloned into the same sites in pJQ200-SK (Quandt and Hynes, 1993), yielding plasmid

113   pJC63. Another 876-bp DNA fragment containing the 3´-*phaC2* gene (120 bp) was PCR amplified

114   using primers JC163-JC164, digested with BamHI-SalI and then cloned into the same sites in pJC63

115   to obtain plasmid pJC69. An omega-Km cassette was obtained from pHP45Ω-Km (Fellay et al.,

116   1987) by BamHI digestion and then inserted into the same site in pJC69 to obtain pJC71. Plasmid

117   pJC71 was then conjugated into *P. putida* PpUW1 in a triparental mating using helper plasmid

118   pRK600. Single cross-over recombination of pJC71 into the *P. putida* chromosome was selected

119   with Rif and Gm. A double cross-over event was achieved by growing a single $Rif^R Gm^R$ colony

120   overnight in LB, making serial dilutions and then spreading on LB Km with 5% sucrose. The

121   resulting PHA^- strain PpUW2 was verified by examining Gm sensitivity (lost plasmid backbone) and

122   absence of PHA production in LB supplemented with 0.5% octanoate (w/v), and was further

123   confirmed by PCR amplification analysis.

124

125   **2.3. Phenotypic screening for *phaC* genes from metagenomic library clones**

126   Construction of the metagenomic DNA library (11AW) of agricultural wheat soil was

127   described previously (Cheng et al., 2014). The 11AW library contains 9×10^6 clones hosted in *E.*

128   *coli* HB101. The pooled library clones (0.5 ml of 300 ml stock) were conjugated *en masse* into *P.*

129   *putida* PpUW2 (PHA^-) with the helper plasmid pRK600. Mating mixture was diluted serially with

130   0.85% NaCl and ~20,000 transconjugants recovered on each LB Tc plate (15 mm × 150 mm)

131   supplemented with 0.5% Na octanoate (w/v). The plates were incubated at 30°C for 24 h and then

132 kept at 22°C for 2-6 days. Potential PHA$^+$ clones of opaque white colour were streak purified on the

133 LB octanoate plates, and verified on 0.1N M63 minimal medium plates (Escapa et al., 2011)

134 supplemented with 0.5% octanoate (w/v) and Nile red (0.5 µg/ml). PHA$^+$ cosmids were transferred

135 from *P. putida* to *E. coli* DH5α(Rif^R), mobilized by pRK600. The cosmid DNA was isolated from *E.*

136 *coli* strains using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), digested by EcoRI-HindIII-

137 BamHI (Thermo Scientific), and then resolved on 1% TAE agarose gels. Cosmids with distinct

138 electrophoretic patterns were conjugated back to *P. putida* PpUW2 (PHA$^-$) to confirm their PHA$^+$

139 phenotype.

140

141 **2.4. Complementation of *S. meliloti* (*phbC*)**

142 11AW cosmid DNA encoding Class I and II *phaC* genes was introduced into *S. meliloti*

143 Rm11476 (Schallmey et al., 2011) via triparental conjugation. Transconjugants were selected on LB

144 SmNmTc plates, and then streaked on YM plates containing Nile red (Schallmey et al., 2011)for

145 visualizing PHB production.

146

147 **2.5. DNA sequencing and bioinformatics**

148 KOD Xtreme DNA polymerase (Novagen) was used for all PCR. Primers are listed in Table

149 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

150 and 68°C for 30 s, and final extension at 68°C for 10 min. The internal regions of *phaC* genes were

151 PCR amplified from PHA$^+$ cosmids using primer phaCF1 and phaCR4 (Sheu et al., 2000). When no

152 PCR product of correct size was obtained, a semi-nested PCR was performed with the primers

153 phaCF2 and phaCR4 as described previously (Sheu et al., 2000). PCR products were resolved on 2%

154 agarose gels, isolated from the gels using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic),

7

155   and then cloned into pJET1.2 vector (Thermo Scientific). Sequences of the cloned partial *phaC*

156   genes were obtained using pJET1.2 sequencing primers (Table 2).

157   End sequences of PHA$^+$ cosmid DNA were obtained by Sanger sequencing using universal

158   M13F and M13R primers. For high throughput sequencing, 68 cosmids of 11AW PHA$^+$ clones and

159   28 cosmids from other research projects were grouped into 24 pools. Tagmentation of pooled cosmid

160   DNA, PCR amplification and clean-up, and library normalization were performed with the Nextera

161   XT DNA library and index kits (Illumina), according to the supplier's recommendation. The library

162   was sequenced using MiSeq 500-cycle version 2 reagent (Illumina). DNA sequences were

163   assembled using SPAdes Genome Assembler 3.5.0 (BaseSpace, Illumina). Cosmid clones were

164   identified based on the available Sanger end sequences and ORFs were annotated with

165   MetaGeneMark (Zhu et al., 2010). Predicted protein sequences were analyzed by BLAST against the

166   non-redundant protein databases (NCBI). Multiple-sequence alignment was performed using

167   MUSCLE (Edgar, 2004). Phylogenetic analysis was performed using MEGA6 (Tamura et al., 2013).

168   Origin of cloned metagenomic DNA was predicted by PhyloPythia (Patil et al., 2012).

169

170   **2.6. Cloning of *phaC1$_{Pp}$* gene of *P. putida* KT2440**

171   To construct a *phaC* expression cosmid for a positive control, the *phaC1$_{Pp}$* gene (*Pp_5003*)

172   of *P. putida* KT2440 was PCR amplified using primers JC159-JC160 and cloned into the HindIII

173   and BamHI sites in the broad-host-range vector pRK7813 (Cheng et al., 2014), to obtain construct

174   pJC67 (P*lac*::*phaC1$_{Pp}$*). A stop codon (TGA) for terminating the translation of *lacZα* upstream of the

175   *phaC1$_{Pp}$*, and a ribosome-binding site (AGGAG) were incorporated into the primer JC159 (Table 2).

176

177   **2.7. Subcloning *phaC* genes of 11AW metagenomic clone 16**

178   The Class II *phaC1$_{16}$* gene of 11AW clone 16 was PCR amplified with oligos JC301 and

179   JC302, and cloned into the HindIII-EcoRI sites in pRK7813 to obtain pJC123. The *phaC2$_{16}$* was

180   amplified using primers JC337 and JC334, and inserted into the EcoRI site in pJC123, yielding

181   pJC157. The orientation of cloned *phaC2$_{16}$* was verified by restriction enzyme mapping. The

182   *phaC2$_{16}$* gene was also obtained by PCR using oligos JC333 and JC334, and cloned into the HindIII-

183   EcoRI sites in pRK7813 to yield pJC159. A stop codon (TGA) for terminating the translation of

184   *lacZα* gene and ribosome-binding sites were added in primers JC333 and JC337. The cloned *phaC*

185   genes were verified by DNA sequencing.

186

187   **2.8. Estimation of PHA production by fluorescent spectrometry**

188   A modification of the previously described method (Wu et al., 2003) was used. *S. meliloti*

189   cells (2.5%, v/v) were added to YM medium. All cultures were incubated at 30°C and 200 rpm for

190   48-72 h. The OD$_{600}$ values of 200 μl cultures were measured in 96-well microtiter plates with the

191   Multiskan Spectrum spectrophotometer (Thermo Labsystems). Samples (180 μl) were stained with

192   20 μl of Nile red (2.5 μg/ml) in the dark for 1 h. Fluorescent intensity was measured at excitation

193   (485 nm) and emission (595 nm) with the FilterMax F5 Multi-mode microplate reader (Molecular

194   Devices). PHA content was calculated based on the equation (fluorescent intensity/OD$_{600}$×0.9).

195

196   **2.9. PHA production, extraction and characterization**

197   *P. putida* PpUW1 (PHA$^+$) or PpUW2 (PHA$^-$) carrying PHA$^+$ clones were grown overnight in

198   LB with or without Tc, washed once with 0.85% NaCl, and then subcultured (2%, v/v) in the 0.1N

199   M63 medium with or without Tc, supplemented with 0.5% Na octanoate, 0.5% nonanoic acid or 1%

200   gluconic acid. The cultures were grown at 30$^°$C and at 200 rpm for 48 hrs. Cells were collected by

201   centrifugation at 20°C and at $9000 \times g$ for 15 min, washed once with deionized water. Cell dry

202   weight (CDW) was obtained after drying the cells at 95°C for 24 hr. For PHA methanolysis, cell

203   pellet (~15 mg) was suspended in 2 ml of chloroform and 2 ml of methanol with 15% H_2SO_4 (v/v),

204   and incubated at 100°C for 5 h. After the reaction mixture was cooled down to 20°C, 1 ml of

205   deionized water was added, and vortexed vigorously for 1 min. The chloroform phase was passed

206   through a cotton plug in a Pasteur glass pipette to remove any cell debris. Methanolyzed sample (1

207   µl) was analyzed with GC-MS (an Agilent 7975B GC equipped with Agilent 5975B inert XL EI/CI

208   MSD and an HP-5MS capillary column). Oven temperature was run at initial 50°C for 5 min with a

209   ramp of 20°C/min to 280°C, and then held for 10 min. The flow rate of Helium carrier gas was 1.2

210   ml/min. Methylated PHA monomers were identified using the Agilent enhanced MSD chemstation

211   (E.02.01.117). PHA standards were kindly provided by Dr. Bruce A. Ramsay (Polyferm Canada).

212

213   **3. Results**

214   **3.1. Isolation of metagenomic clones for PHA production**

215   To employ *P. putida* as a surrogate host to isolate 11AW (agricultural wheat soil)

216   metagenomic library clones encoding functional PhaC, it was first necessary to construct a PHA

217   synthesis mutant. Expression of the contiguous three genes encoding PHA synthase PhaC1$_{Pp}$

218   (Pp_5003), PHA depolymerase PhaZ$_{Pp}$ (Pp_5004) and PHA synthase PhaC2$_{Pp}$ (Pp_5005) in *P.*

219   *putida* KT2440 Rif^{R} derivative PpUW1 was disrupted by deletion of 3437-bp comprised of the 3′

220   region (911 bp) of *phaC1$_{Pp}$* (1677 bp), *phaZ$_{Pp}$* and the 5′ region (1560 bp) of *phaC2$_{Pp}$* (1680 bp) and

221   replacement with an omega-Km kanamycin resistance insert, resulting in the PHA^{-} strain PpUW2.

222   Transfer of the negative control cosmid vector pJC8 to *P. putida* PpUW2 did not result in detectable

223   PHA production. However, introduction of pJC67 (P*lac*::*phaC1$_{Pp}$*) *in trans* restored the PHA^{+}

224   phenotype (Fig. 1). These data suggested that the strain PpUW2 could be used for screening of

225   PhaC-encoding metagenomic clones. The 11AW library clones (Cheng et al., 2014) were transferred

226   to *P. putida* PpUW2 via en masse triparental conjugation. Following selection of *P. putida* PpUW2

227   transconjugants (~4 million) on LB Km Tc plates supplemented with 0.5% Na octanoate, we

228   obtained 72 clones that exhibited greater opacity than the PpUW2 PHA⁻ recipient strain. The PHA-

229   producing phenotype of those clones was verified by visualizing the fluorescence of Nile red-stained

230   PHA in 0.1N M63 minimal medium (Escapa et al., 2011) supplemented with Na octanoate as the

231   sole carbon source (Fig. 1). Restriction digest of DNA from the 72 clones demonstrated 68 distinct

232   restriction patterns, suggesting the presence of a broad diversity of DNA origin in those PHA^{+} clones.

233

234   **3.2. Identification of *phaC* genes encoding PHA synthases**

235   Internal regions (~500 bp) of the *phaC* genes of 18 distinct cosmid clones were initially

236   obtained by PCR amplification using degenerate primers PhaCF1 and PhaCR4 (Sheu et al., 2000).

237   For additional 4 clones, *phaC* fragments (PhaC$_3$, PhaC$_7$, PhaC$_{10}$, and PhaC$_{15}$) of ~400 bp were

238   generated by nested-PCR using primer pair PhaCF2-PhaCR4 as described previously (Sheu et al.,

239   2000). BLASTP analysis of the cloned gene products indicated that 13 PhaC proteins could be

240   grouped into Class I while the other 9 were categorized as Class II PHA synthases.

241   To identify all *pha* genes on the isolated clones, DNA sequences were obtained by high

242   throughput sequencing, and additional *pha* gene loci were identified in partially and fully assembled

243   clones (Table 3; Supplementary 1). The identified PHA synthases could be classified into 3 groups

244   based on amino acid sequences (Fig. 2). Class I *phaC* genes were annotated in 17 clones. The *phaC*

245   and *phaAB* genes were adjacently located in 12 of these clones, whereas the *phaB* gene was located

246   distantly downstream of *phaCA* genes in clone 25 and clone P1N3 (Supplementary 1). A *phaR* gene

11

247 was located immediately downstream of *phaCB* genes in partially assembled clone P11N2. The

248 *phaC* and *phaB* genes flanked a *phaZ* gene in the partially assembled clone P2N8. The metagenomic

249 DNA in these Class I clones was predicted to originate from *Gemmatimonas*, *α-Proteobacteria*

250 (*Sphingomonadaceae*), *β-Proteobacteria* (*Leptothrix, Rubrivivax, Janthinobacterium*) and γ-

251 *Proteobacteria* (*Xanthomonadaceae*) (Table 3).

252   Class II PHA genes are commonly clustered with *phaZ* flanked by two PHA synthase

253 encoding genes, *phaC1* and *phaC2*. Nine clones carried Class II *phaC* genes (Table 3). The *pha* gene

254 locus ($phaC1_1$-*phaZ*-$phaC2_1$-*phaD*-*phaF*-*phaI*) in clone 1 was similar to the canonical locus in

255 *Pseudomonas* including *P. putida* KT2440 (de Eugenio et al., 2010), except the presence of *orf7*

256 encoding a PHA granule associated protein (Pfam09650) (Fig. 3A). Clone 2 and clone 15 were

257 identical to the clone 1, except that they contained an insertion of the 17,013-bp transposon Tn4652

258 at 13,740 nt and 16,362 nt respectively (Supplementary 2). The Tn4652 duplicated target sequences

259 were AACTC in clone 2 and TAGGA in clone 15. The transposon insertions in these clones are in

260 opposite orientation. The same Tn4652 is present in *P. putida* KT2440 genome (3,366,550 -

261 3,383,562 nt, GenBank: AE015451), located distant from the *pha* gene operon. These insertions

262 likely occurred following introduction of the cosmid clones into *P. putida*.

263   Class II clones 5, 16, 20 and 50 had no genes encoding *phaD* or *phaFI* homologs. A *phaJ*

264 gene encoding R-specific enoyl-CoA hydratase was identified in those clones (Fig. 3A;

265 Supplementary 1). The *phaJ* gene was located immediately upstream of the *phaC* genes in clones 16,

266 20 and 50, but downstream of *phaC* in clone 5. In addition, a gene encoding PHA depolymerase

267 (PhaZ) was located downstream of the *phaC* genes in clones 16, 20 and 50. ORFs downstream of the

268 *phaZ* genes in clones 16 and 50 encoded proteins ($PhaC2_{16}$ and $PhaC2_{50}$) homologous to unclassified

12

269   PHA synthases (Fig. 2; Supplementary 1). The proteins shared conserved regions in PhaC enzymes,

270   but were ~30% shorter in N-terminal regions than in those of Class I/II PhaC (Supplementary 2).

271   Two ORFs encoding putative PhaC and PhaE proteins were identified in partially assembled

272   clone P2N10 (Table 3; Supplementary 1). BLASTP showed the P2N10 PhaC protein best matched

273   to the Class III PhaC [GenBank: WP_002708071] (43% identity) from γ-*Proteobacteria Thiothrix*

274   *nivea* (Table 3). The available C-terminal sequence of P2N10 PhaE (234 amino acids) only exhibited

275   26% identity to the Class III PhaE [GenBank: WP_002708072] of *T. nivea*. In addition, the P2N10

276   PhaC protein was only 29% identical to the Class IV PhaC [GenBank: AAD05620] of *Bacillus*

277   *megaterium*. Phylogenetic analysis suggested that P2N10 PHA synthase was distant from Class

278   III/IV PhaC and might represent a new subclass of PHA synthases (Fig. 2).

279   Multiple alignment of amino acid sequences of each of the identified 11AW PhaC proteins

280   showed the conserved catalytic triad (C287, D445 and H473 in $PhaC1_{16}$), a tryptophan essential for

281   dimerization (W391 in $PhaC1_{16}$), and the lipase box GXCXGG (Jia et al., 2000) except that Ala

282   replaced the first Gly residue in $PhaC_{20}$ and second Gly in clone P8N7 (Fig. 3B; Supplementary 3).

283   The serine residue (S229 in $PhaC1_{16}$) required for PhaC activity (Hoppensack et al., 1999) was also

284   conserved in all 11AW PhaC proteins. We chose the PhaC proteins in the Class I clones 14, 18 and

285   25, and the Class II clones 1, 16 and 20, for further study. We examined the nature of PHA produced

286   using different carbon sources that result in substrate production either through fatty acid synthesis

287   or β-oxidation.

288

289   **3.3. Class II clone 1 synthesizes MCL PHA**

290   The cloned metagenomic DNA in clone 1 contained 42,747 bp (Table 3) [GenBank:

291   KT944254] and most likely originated from γ-*Proteobacteria Pseudomonas*. The $PhaC1_1$, PhaZ,

13

292   PhaC2$_1$ and PhaD proteins [GenBank: ALV86243, ALV86281, ALV86274 and ALV86275] were

293   82%, 92%, 73% and 78% identical to the corresponding orthologs (PhaC1, PhaZ, PhaC2 and PhaD)

294   of *P. putida* KT2440. Clone 1 PhaC2$_1$ was phylogenetically related to the PHA synthase 2

295   [GenBank: BAA36202] of *Pseudomonas* sp. 61-3 (Fig. 2). In contrast to other Class II PhaC, the

296   conserved amino acid at position 129 of PhaC2$_1$ was Asp rather than Glu (Fig. 3B). It has been

297   previously demonstrated that substitution of the conserved Glu with Asp in Class II PhaC improves

298   PHA yield with an increase of 3-hydroxybutyrate monomer (Matsumoto et al., 2005).

299   PHA synthesis in clone 1 was likely contributed solely by PhaC2$_1$ because the 5'-region of

300   *phaC1$_1$* was absent in the cloned DNA. Expression of the functional *phaZ* and *phaC2$_1$* might be

301   driven by the promoters upstream of the individual genes, as occurs in *P. putida* KT2440 (de

302   Eugenio et al., 2010). *P. putida* PpUW2 (PHA$^-$) carrying clone 1 synthesized 3-hydroxyhexanote

303   (3HHx, C6) and 3-hydroxyoctanoate (3HO, C8) copolymer of ~95% C8 monomer when grown with

304   octanoate, similar to the PHA produced by wild-type *P. putida* PpUW1 (Table 4). When PpUW2

305   (clone 1) was grown with nonanoic acid, 3-hydroxynonanoate (3HN, C9) and 3-hydroxyheptanoate

306   (3HP, C7) were incorporated into the PHA with greater 3HP than that synthesized by wild type

307   PpUW1 (Table 4). These data suggest that those monomers were derived from β-oxidation of

308   octanoate and nonanoic acid.

309   When gluconic acid was used as the sole carbon source, 3HO and 3-hydroxyldecanoate

310   (3HD, C10) copolymer was produced in PpUW2 (PHA$^-$) carrying clone 1, very similar to the

311   production of the parental strain PpUW1 (PHA$^+$) (Table 4). The monomer composition of PHA was

312   similar in both strains, but total PHA accumulation was ~4 fold higher in gluconic acid-grown

313   PpUW2 with clone 1 (Table 4). These results suggest that Class II PhaC1$_2$ was able to synthesize

314   MCL (C6-C10) PHA.

14

315

3.4. Class II clone 16 synthesizes SCL-MCL copolymer

316

317   Clone 16 contains 33,900-bp metagenomic DNA [GenBank: KT944263] probably originated

318   from *β-Proteobacteria Methylibium* (Table 3). A *phaJ* gene (*phaJ16*) is predicted to encode a R-

319   specific enoyl-CoA hydratase [GenBank: ALV86416] (Fig. 3A), which converts the β-oxidation

320   intermediate 2-enoyl-CoA to (R)-3-hydroxyacyl-CoA used as PhaC substrate (Fukui et al., 1998).

321   The PhaJ$_{16}$ was phylogenetically related to the PhaJ4a$_{Re}$ and PhaJ4b$_{Re}$ of *Ralstonia eutropha*

322   (Kawashima et al., 2012), and PhaJ4$_{Pp}$ of *P. putida* (Sato et al., 2011) (Supplementary 4A). The

323   amino acid residues Asp39 and His44 conserved at the active site of the dehydratases were identified

324   in PhaJ$_{16}$ (Supplementary 3B), except that the Ser residue was replaced by Pro71 in PhaJ$_{16}$, same as

325   in the PhaJ orthologs of *R. eutropha* and *P. putida* PhaJ4.

326   The three ORFs downstream of the *phaJ16* gene encoded PhaC1$_{16}$, PhaZ and PhaC2$_{16}$ (Fig.

327   3A). PhaC1$_{16}$ of Class II PHA synthase most closely matched the *Pseudomonas* sp. F15 PhaC1

328   [GenBank: WP_021487788], exhibiting 48% amino acid identity. PhaC1$_{16}$ was also 42.5% and

329   43.4% identical to the PhaC1 and PhaC2 proteins of *P. putida* KT2440. Thr318 and Val478 in the

330   PhaC1$_{16}$ might favour the 3HB incorporation into PHA as occurred in the engineered PhaC proteins

331   (Chen et al., 2014; Takase et al., 2003). Another putative PHA synthase encoded by the *phaC216*

332   gene downstream of the *phaC116 phaZ* genes was homologous to unclassified PHA synthases (Fig.

333   2).

334   The quantity of PHA accumulated in *P. putida* UW2 carrying clone 16 was comparable to that

335   of parental PpUW1 when grown with octanoate, nonanoic acid or gluconic acid (Table 4). In

336   contrast to the MCL PHA (C6-C8) present in octanoate-grown PpUW2 (clone 1) or wild-type

337   PpUW1, about 20% 3HB was detected in the PHA (C4-C6-C8) accumulated in PpUW2 carrying

15

338   clone 16 under the same conditions (Table 4). When PpUW2 (clone 16) was grown with nonanoic

339   acid (C9), 3-hydroxyheptanoate (3HP, C7), 3-hydroxyvalerate (3HV, C5) and 3HB (C4) were

340   polymerized, which was in contrast to the 3HP-3HN PHA in nonanoic acid-grown PpUW with

341   PhaC1 or wild type PpUW1. Additionally, equal amounts of C6 and C8 monomers were detected in

342   the PHA from gluconic acid-grown PpUW2 with clone 16. However, PHA with C8 and C10

343   monomers was accumulated in PpUW2 (clone 1) and PpUW1 under the same growth conditions.

344   These data implied that the PhaC1$_{16}$ and/or PhaC2$_{16}$ prefer substrates with shorter carbon chains than

345   the PhaC2$_1$ in clone 1 and both PhaC proteins of *P. putida* KT2440.

346   To further elucidate the activity of PhaC1$_{16}$ and PhaC2$_{16}$, the genes were cloned

347   downstream of the constitutively active P*lac* promoter. When *P. putida* PpUW2 carrying pJC123

348   (P*lac*::*phaC1$_{16}$*) was grown with octanoate, SCL-MCL copolymer was of similar composition of C4-

349   C6-C8 monomers to that synthesized by clone 16 (Table 4). However, PHA was not detected in

350   strain PpUW2 carrying pJC159 (P*lac*::*phaC2$_{16}$*) under the same growth conditions. Coexpression of

351   *phaC1$_{16}$* and *phaC2$_{16}$* in pJC157 (P*lac*::*phaC1$_{16}$-phaC2$_{16}$*) in octanoate-grown PpUW2 resulted in

352   synthesis of PHA with the similar monomer composition as those in the polymers produced in

353   PpUW2 carrying clone 16 or pJC123 (P*lac*::*phaC1$_{16}$*) (Table 4). These data indicated that only the

354   PhaC1$_{16}$ was involved in SCL-MCL PHA biosynthesis in clone 16.

355   The *pha* gene locus of clone 16 was very similar to that of clone 50 (Supplementary 1), but

356   the DNA in clone 50 probably originated from γ-*Proteobacteria Hydrocarboniphata* (Table 3). Both

357   the Class II PhaC1 and unclassified PhaC2 in clones 16 and 50 shared 84% identity respectively.

358   Expression of the *pha* genes was probably driven by a promoter upstream of *phaJ* genes.

359

360   **3.5. Class II clone 20 synthesizes MCL or SCL-MCL PHA**

16

361   PHA$^+$ clone 20 contained 30,964-bp metagenomic DNA [GenBank: KT944266], which

362   probably originated from *Salinibacter* within the *Bacteroidetes* (Table 3). The cloned DNA

363   contained only a single *phaC$_{20}$* gene (Fig. 3A). A putative fatty acid CoA ligase/ synthetase

364   [GenBank: ALV86493] (ORF9; Supplementary 1) might be involved in degradation of fatty acids to

365   acyl-CoA (the first step of β-oxidation cycle). The acyl dehydratase PhaJ$_{20}$ [GenBank: ALV86492]

366   was 56% identical to the PhaJ$_{16}$. Conserved amino acid residues Asp41 and His46 at the active sites

367   of the acyl dehydratases were present (Supplementary 4B) except that the Ser residue was replaced

368   by Pro73 in the PhaJ$_{20}$ protein. The *phaJ$_{20}$*, *phaC$_{20}$* and *phaZ* genes in clone 20 were similar to those

369   of the corresponding genes in clones 16 and 50 (Fig. 3A; Supplementary 1). In addition, the *pha*

370   gene cluster of clone 20 was similar to that in *Burkholderia* sp. 383 (1,576,365 -1,581,635 nt)

371   [GenBank: NC_007511], but a gene encoding a phasin protein is located between the *phaC* and

372   *phaZ* genes in the *Burkholderia* genome.

373   The single PhaC$_{20}$ [GenBank: ALV86493] best matched to its ortholog from *Burkholderia lata*

374   at 59% amino acid sequence identity [GenBank: WP_011354994]. PhaC$_{20}$ was 46-53% identical to

375   the PhaC1$_{16}$, PhaC2 and PhaC5 as well as PhaC1 and PhaC2 of *P. putida* KT2440. Presence of amino

376   acid residues Thr353 and Val516 in PhaC$_{20}$ (Fig. 3B) might favour substrates with fewer carbons

377   than those preferred by clone 1 and *P. putida* PhaC proteins, as previously demonstrated with

378   engineered Class II PHA synthases (Chen et al., 2014; Takase et al., 2003).

379   Clone 20 was able to synthesize 3HHx-3HO copolymer (C6-C8) with octanoate (Table 4), as

380   occurred in PpUW2 (clone 1) and *P. putida* wild type. However, the proportion of C6 monomer

381   increased 4-35 fold in the PHA isolated from PpUW2 (clone 20), compared to that in PpUW2 (clone

382   1) and wild-type PpUW1 (Table 4). When nonanoic acid was supplied, PpUW2 (clone 20) produced

383   PHA with 8% 3-hydroxyvalerate (3HV), which was absent in the PHA synthesized by clone 1 and

17

384   PpUW1 (Table 4). In contrast to the SCL-MCL PHA produced by clone 16, 3HB was not detected in

385   the nonanoic acid or octanoate-grown PpUW2 with clone 20.

386   In contrast to the absence of detectable C6 monomer in the PHA produced in PpUW2 with

387   clone 1 or *P. putida* wild type, 3HHx (~10%) was incorporated into the MCL PHA (C6-C8-C10) in

388   PpUW2 (clone 20) (Table 4) when gluconic acid was used as the sole carbon source. About half the

389   quantity of monomers was C10 in the PHA accumulated in PpUW2 (clone 20) grown with gluconic

390   acid. These results suggested that the carbon chain lengths of favourite monomers of $PhaC_{20}$ were

391   between those of $PhaC1_{16}$ and $PhaC2_{1}$.

392

393   **3.6. Class I clone 14 synthesizes SCL-MCL PHA**

394   The DNA in clone 14 probably originated from *β-Proteobacteria Methylibium* (Table 3). The

395    $PhaC_{14}$, PhaA and PhaB [GenBank: ALV86397, ALV86398 and ALV86399] were 65%, 76% and

396   74% identical to the PhbC1 (A16_1437), PhbA (A16_1438) and PhbB1 (A16_1439) of *C. necator*

397   H16, respectively. In addition, $PhaC_{14}$ was 48% identical to the PhbC protein (Smc002960) of *S.*

398   *meliloti* Rm1021. The best match of $PhaC_{14}$ was the PhaC protein of *Burkholderia* JOSHI_001 at

399   74% (Table 3). The *phaC$_{14}$AB* locus of clone 14 is similar to that of clones 8, 21 and 23

400   (Supplementary 1). Expression of the *phaC$_{14}$*, *phaA* and *phaB* genes in clone 14 was most likely

401   driven by a promoter upstream of the *orf9* encoding a hypothetical protein.

402   When PpUW2 (clone 14) was grown with nonanoic acid, 7% 3HB was incorporated into

403   3HB-3HV copolymer (Table 4). 3HB was the primary monomer of PHA in octanoate- or gluconic

404   acid-grown PpUW2. Both C6 and C8 monomer were present in the PHA with gluconic acid as

405   carbon source. However, C8 monomer was absent in the PHA when octanoate was supplied (Table

406   4). *P. putida* PpUW2 carrying clone 14 produced the highest quantity of PHA among all the PHA[+]

18

407   clones (Table 4). These results indicated that PhaC$_{14}$ could synthesize SCL-MCL copolymer though

408   3HB was the dominant monomer.

409

410   **3.7. Class I clone 18 synthesizes SCL and SCL-MCL PHA**

411   Clone 18 contained 37,818-bp metagenomic DNA [GenBank: KT944264], which probably

412   originated from *β-Proteobacteria Rubrivirax* (Table 3). The annotated *phaC$_{18}$*, *phaA* and *phaB* were

413   organized in one operon (Fig. 3A), as also occurred in clones 27 and 51 (Supplementary 1). PhaC$_{18}$

414   [GenBank: ALV86462] shared 100% and 99% identity to PhaC$_{27}$ [GenBank: ALV86602] and

415   PhaC$_{51}$ [GenBank: ALV86651] respectively. The PhaC$_{18}$ protein best matched to its homolog in

416   *Azohydromonas austrilica* (Table 3), and was 71%, 49% and 61% identical to the PhaC$_{14}$, PhbC$_{Sm}$

417   and PhbC1$_{Re}$ respectively.

418   SCL-MCL PHA (C4-C6-C8) was produced when PpUW2 (clone 18) was grown with

419   octanoate (Table 4), as occurred in PpUW2 with clone 16. However, the fraction of C6-C8

420   monomers was only 3% in the PHA from clone 18, in contrast to the ~83% in the copolymer from

421   clone 16. The yield of PHA was ~7 fold less than that produced by clone 14, but C8 monomer was

422   absent (Table 4). PHA of 3HB-3HV monomers was accumulated in nonanoic acid-grown PpUW2

423   (clone 18). The ratio of 3HB to 3HV increased 5 fold compared to that in PpUW2 (clone 14) under

424   the same growth conditions (Table 4). When PpUW2 carrying clone 18 was grown on gluconic acid,

425   only PHB was produced. The yield of polymer increased 5 fold compared to that in octanoate-grown

426   cells, and also the second highest quantity of PHA produced among all strains grown on gluconic

427   acid. These results implied that PhaC$_{18}$ could synthesize SCL- and SCL-MCL PHA dependent of the

428   available carbon source.

429

430 **3.8. Class I clone 19 synthesizes SCL and SCL-MCL PHA**

431   Clone 19 contained 25,085-bp of DNA that might originate from *β-Protoebacteria Leptothrix*

432 (Table 3) [GenBank: KT944265]. Three ORFs encoded $PhaC_{19}$, PhaA and PhaB for PHA

433 biosynthesis (Fig. 3A). $PhaC_{19}$ shared 99% identity with $PhaC_6$ and $PhaC_3$ (Fig. 2). The best match

434 for $PhaC_{19}$ is the ortholog of *A. austrilica* at 73% (Table 3). The $phaC_{19}AB$ genes in clone 19

435 probably consisted of one operon, and the promoter was located upstream of the $phaC_{19}$.

436   Clone 19 was able to synthesize SCL-MCL PHA (C4-C6-C8) with octanoate but only PHB

437 (C4) in gluconic acid-grown PpUW2 (Table 4), as occurred in PpUW2 carrying clone 18. However

438 ~5 fold more PHA was synthesized by clone 19 than was produced by clone 18 in octanoate medium

439 (Table 4). In contrast, the quantity of PHA was lower in gluconic acid-grown PpUW2 with clone 19

440 than it was in the same host carrying clone 18. Moreover, PHA composed of 3HB-3HV monomers

441 was synthesized in noanoic acid-grown PpUW2 (clone 19), as occurred in PpUW2 (clone 18). These

442 data suggested that $PhaC_{19}$ was able to synthesize both SCL and SCL-MCL PHA.

443

444 **3.9. Class I clone 25 synthesizes SCL and SCL-MCL PHA**

445   Clone 25 contained 41,012-bp of DNA [GenBank: KT944269], which was predicted to

446 originate from *β-Proteobacteria Variovorax* (Table 3)*. The $phaC_{25}$ and *phaA* genes were apparently

447 located in a single operon (Fig. 3A). A *phaB* gene was located 9 open reading frames downstream of

448 the *phaA*. The $phaC_{25}A$ and *phaB* loci in clone 25 are similar to those in clone P1N3 (Supplementary

449 1), though the *phaB* gene in clone P1N3 was four open reading frames away from $phaC_{P1N3}A$. PhaC

450 proteins of clones P1N3 and 25 shared 77% identity. In addition, $PhaC_{25}$ was 35% and 61% identical

451 to $PhbC_{Sm}$ and $PhbC1_{Re}$. The conserved amino acid Asp130 affecting PhaC substrate specificity

452 (Matsumoto et al., 2005) was replaced by Ala152 (Fig. 3B).

20

453   SCL-MCL PHA (C4-C6-C8) was produced in PpUW2 harbouring clone 25 when grown on

454   octanoate, and only PHB (C4) was isolated from gluconic acid-grown cells (Table 4). Similar results

455   were obtained in PpUW2 with clone 18 or clone 19 under the same conditions. However, the

456   quantity of C6-C8 monomers was greater with clone 25 that those with clone 18 or 19. Nonanoic

457   acid-grown PpUW2 (clone 25) produced 3HB-3HV copolymer similar to that from PpUW2 with

458   clone 18 or clone 19 (Table 4). These data suggested that $PhaC_{25}$ had similar properties to those of

459      $PhaC_{18}$ and $PhaC_{29}$, but different from that of $PhaC_{14}$.

460

461 **3.10. Complementation of *S. meliloti phaC* mutant with 11AW PHA[+] clones**

462   In an earlier study, *S. meliloti* Rm11476 (*exoY*::Tn5 *phaC*::Tn5-233) had been employed for

463   functionally harvesting metagenomic library clones pCX92, pCX9M1, pCX9M3 and pCX9M5

464   expressing Class I PHA synthases (Schallmey et al., 2011). The PhaA/PhbB and PhaB/PhbB

465   proteins encoded by the cloned metagenomic DNA and/or *S. meliloti* genome could supply (R)-3-

466   hydroxybutyrate for PHB biosynthesis catalyzed by the PhaC/PhbC enzymes. In order to examine

467   the possibility that the PHA[+] clones of 11AW metagenomic library were able to complement the

468   PHA[-] phenotype of *S. meliloti* Rm11476, The Class I clones 14, 18, 19 and 25 (Table 3) were

469   introduced into *S. meliloti* Rm11476. In each case, complementation was observed, indicating PhaC

470   function in the *S. meliloti* background (Fig. 4), with greatest accumulation of PHA observed with

471   clone 14. However, when clones pCX92, pCX9M1, pCX9M3 and pCX9M5 were transferred to *P.*

472   *putida* PpUW2 (PHA[-]), PHA was not detected in the recombinant strains grown with either

473   octanoate or gluconic acid (data not shown). These results strengthen the argument that multiple

474   hosts should be employed to screen for novel *phaC* genes.

475   We then asked whether Class II *phaC* could function in the *S. meliloti* background. First, we

476   observed no detection of PHA when *phaC1$_{Pp}$* (P*lac*::*phaC1$_{Pp}$*) was expressed in Rm11476 even

477   though the *lac* promoter is functional and 3-hydroxybutyrate is supplied by the PhbAB$_{Sm}$. Similarly,

478   PHA$^+$ clones 1 and 20 carrying Class II *phaC* genes also failed to complement the PHB$^-$ phenotype

479   of *S. meliloti* Rm11476 (Fig. 4), implying that the *phaC2$_1$* and *phaC$_{20}$* genes were not expressed

480   and/or the substrate (R)-3-hydroxyalkanoic acids were absent in *S. meliloti*. However, when clone 16

481   was introduced into *S. meliloti* Rm111476, PHA was accumulated to a similar level as with

482   Rm11476 containing PHB$^+$ clones pCX92, pCX9M1, pCX9M3 or pCX9M5 (Fig. 4). These data

483   indicated that the *phaC1$_{16}$* gene was expressed and its product was functional. These results suggest

484   that *P. putida* is a more permissive surrogate host than *S. meliloti* for screening of novel PHA

485   synthases.

486

487   **4. Discussion**

488   Biosynthesis of SCL-MCL PHA requires PHA synthases having a broad range of substrate

489   specificity and SCL and MCL precursors (R-3-hydroxyacyl-CoAs) must be available within the cell.

490   Building on earlier work that has sought to mine metagenomic DNA for modification of bacterial

491   PHA production, we have successfully isolated cosmid clones that are able to functionally

492   complement a *P. putida* PHA synthesis mutant. DNA sequence analysis revealed that the isolated

493   metagenomic DNA originated from a broad diversity of bacteria, and encoded either Class I or Class

494   II PHA synthases enzymes. Of note is that DNA from most of the isolated clones did not very

495   closely match known sequences. Also of interest is the influence that the clones had on the quality

496   and quantity of PHA produced in the *P. putida* surrogate host background.

22

497   A major appeal of PHA is that, as alternatives to conventional fossil fuel-derived plastics,

498   they can be produced from renewable resources. The potential to improve the efficiency and cost of

499   production, and expand the range of polymers and copolymers that are available for production,

500   should impact on competitiveness and adoption of these materials in the marketplace. In this context,

501   it is important to understand how the properties of PHA are determined by the PHA synthase

502   enzyme and the substrate that is available to those enzymes, which is in turn influenced by the

503   culture conditions and the metabolic pathways leading to substrate formation (Meng et al., 2014).

504   PHA consisting of both short-chain-length (SCL, ≤C5) and medium-chain-length (MCL, ≥C6)

505   monomers have great flexibility, decreased breakage and reduced melting point (Noda et al., 2005).

506   Most naturally occurring PhaC proteins polymerize either SCL- or MCL-monomers, but there are a

507   few know examples such as the Class I PhaC of *Aeromonas caviae* (Fukui and Doi, 1997) and some

508   *Pseudomonas* Class II PhaC (Matsusaki et al., 1998) that can synthesize SCL-MCL copolymers.

509   In the present work, heterologous complementation of the *P. putida* PHA⁻ strain facilitated

510   efficient simultaneous screening of millions of 11AW metagenomic clones, resulting in recovery of

511   a greater number of novel Class I, Class II, and unclassified PHA synthases than from previous

512   reports (Cheema et al., 2012; Schallmey et al., 2011). Additionally, the clones carrying Class I *phaC*

513   isolated in *P. putida* were able to complement the PHA⁻ phenotype of *S. meliloti*, but the previously

514   isolated Class I clones isolated in *S. meliloti* (Schallmey et al., 2011) failed to synthesize PHA in *P.*

515   *putida.* These results support the development of multi-host systems to increase the chances of the

516   successful expression of *pha* genes of interest.

517   Biosynthesis of SCL-MCL PHA of various forms by Class I clones 14, 18, 19 and 25 and

518   Class II clones 16 and 20 suggested availability of a full range of R-3-hydroxyacyl-CoAs derived

519   from β-oxidation of fatty acids such as nonanoic acid and octanoate. The differential monomer

23

520   compositions in PHA synthesized by the Class II enzymes $PhaC_{16}$ and $PhaC_{20}$ are of particular

521   interest, especially since most Class II PHA synthases are unable to incorporate SCL monomers such

522   as 3HB and 3HV. Changes of conserved amino acid residues Ser325Thr, Leu484Val, and

523   Glu508Leu (conserved positions in *P. putida* PhaC1) have been identified to be involved in PhaC

524   substrate specificity and PHA yield (Chen et al., 2014; Shozui et al., 2010; Takase et al., 2003).

525   Presence of the amino acids Thr353 (corresponding to *P. putida* PhaC1 position 325), Val516

526   (corresponding to *P. putida* PhaC1 position 484) and Ala540 (corresponding to *P. putida* PhaC1

527   position 508) $PhaC_{20}$ might contribute to the enzyme's ability to incorporate SCL monomers

528   compared to those of 11AW $PhaC2_1$ and PHA synthases of *P putida* KT2440. Similarly, the

529   presence of three residues (Thr317, Val478 and Leu502) at these positions in $PhaC1_{16}$, most likely

530   resulted in increased 3HB content as demonstrated previously with engineered Class II PHA

531   synthases (Chen et al., 2014; Shozui et al., 2010; Takase et al., 2003).

532

533   **5. Conclusions**

534   We obtained 27 clones encoding Class I, II and unclassified PHA synthases by functional

535   screening of soil metagenomic cosmid clones in a *P. putida* PHA⁻ strain. Seven clones that were

536   characterized in more detail were able to produce a broad range of polymers and copolymers,

537   including SCL-MCL mixtures, depending on carbon source. Through this work we have

538   demonstrated the potential for using metagenome-derived clones for production of a variety of PHAs

539   of possible industrial utility. The collection of PHA metabolism genes from uncultivated organisms

540   provides not only a resource for production strain development, but also a series of sequence

541   templates that could prove useful in enzyme engineering efforts directed towards generation of PHA

542   products with desired properties.

543

Nucleotide sequence accession numbers

545   Complete sequences of 11AW metagenomic PHA$^+$ DNA have been deposited in GenBank

546   nucleotide sequence database [GenBank: KT944254-KT944278, KU728995 and KU728996].

547

Abbreviations

549   *P. putida*, *Pseudomonas putida*; PHA, polyhydroxyalkanoate; PhaA/PhbA, β-ketothiolase;

550   PhaB/PhbB, acetoacetyl-CoA reductase; PhaC/PhbC, PHA synthase; PhaZ, PHA depolymerase;

551   PhaJ, R-specific enoyl-CoA hydratase; 3HA, 3-hydroxyalkanoate; 3HB, 3-hydroxybutyrate; 3HV.

552   3-hydroxyvalerate; 3HHx, 3-hydroxyhexanoate; 3HP, 3-hydroxyheptanoate; 3HO, 3-

553   hydroxyoctanoate; 3HN, 3-hydroxynonanoate; 3HD, 3-hydroxydecanoate; SCL, short chain length;

554   MCL, medium chain length; *S. meliloti*, *Sinorhizobium meliloti*; GC-MS, gas chromatography-mass

555   spectrometry; CDW, cell dry weight.

556

Competing interests

558   The authors declare that they do not have competing interests.

559

Authors's contributions

561   JC and TCC conceived and designed this study, carried out all the experiments, and prepared

562   the manuscript. Both authors read and approved the final manuscript.

563

564

565

25

Acknowledgements

566

567   This work was financially supported by a New Directions grant from Ontario Ministry of

568   Agriculture, Food, and Rural Affairs (OMAFRA), and partially by a Strategic Projects grant from

569   the Natural Sciences and Engineering Research Council of Canada (NSERC) and by Genome

570   Canada for the project "Microbial Genomics for Biofuels and Co-Products from Biorefining

571   Processes". We are very grateful to Ricardo Nordeste for assisting with GC-MS analysis of PHA. We

572   wish to thank Dr. Bruce Ramsay (Polyferm Canada) for providing PHA standards, and Dr. Josh

573   Neufeld for use of the FilterMax F5 Multi-mode microplate reader.

Supplementary

574

575   Supplementary 1: *pha* gene locus in 11AW PHA[+] metagenomic clones

576   Supplementary 2. Transposon Tn4652 in 11AW Clones 2 and 15

577   Supplementary 3: Multiple sequence alignment of polyhydroxyalkanoate synthases of 11AW PHA[+]

578   metagenomic clones

579   Supplementary 4: Phylogenetic analysis and sequence alignment of R-specific enoyl-CoA hydratase

580   PhaJ

581
582   **References**
583
584   Cheema, S., Bassas-Galia, M., Sarma, P.M., Lal, B., Arias, S., 2012. Exploiting metagenomic
585   diversity for novel polyhydroxyalkanoate synthases: Production of a terpolymer poly(3-
586   hydroxybutyrate-co-3-hydroxyhexanoate-co-3-hydroxyoctanoate) with a recombinant
587   *Pseudomonas putida* strain. Bioresource Technology 103, 322–328.
588   doi:10.1016/j.biortech.2011.09.098
589   Chen, Y.-J., Tsai, P.-C., Hsu, C.-H., Lee, C.-Y., 2014. Critical residues of class II PHA synthase for
590   expanding the substrate specificity and enhancing the biosynthesis of polyhydroxyalkanoate.
591   Enzyme and Microbial Technology 56, 60–66. doi:10.1016/j.enzmictec.2014.01.005
592   Cheng, J., Pinnell, L., Engel, K., Neufeld, J.D., Charles, T.C., 2014. Versatile broad-host-range
593   cosmids for construction of high quality metagenomic libraries. Journal of Microbiological
594   Methods 99, 27–34. doi:10.1016/j.mimet.2014.01.015
595   de Eugenio, L.I., Escapa, I.F., Morales, V., Dinjaski, N., Galàn, B., García, J.L., Prieto, M.A., 2010.
596   The turnover of medium-chain-length polyhydroxyalkanoates in *Pseudomonas putida* KT2442

26

597   and the fundamental role of PhaZ depolymerase for the metabolic balance. Environ. Microbiol.
598   12, 207–221. doi:10.1111/j.1462-2920.2009.02061.x
599   Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
600   complexity. BMC Bioinformatics 5, 113–19. doi:10.1186/1471-2105-5-113
601   Escapa, I.F., Morales, V., Martino, V.P., Pollet, E., Avérous, L., García, J.L., Prieto, M.A., 2011.
602   Disruption of β-oxidation pathway in *Pseudomonas putida* KT2442 to produce new
603   functionalized PHAs with thioester groups. Appl. Microbiol. Biotechnol. 89, 1583–1598.
604   doi:10.1007/s00253-011-3099-4
605   Fellay, R., Frey, J., Krisch, H., 1987. Interposon mutagenesis of soil and water bacteria: a family of
606   DNA fragments designed for in vitro insertional mutagenesis of gram-negative bacteria. Gene
607   52, 147–154.
608   Foong, C.P., Lau, N.-S., Deguchi, S., Toyofuku, T., Taylor, T.D., Sudesh, K., Matsui, M., 2014.
609   Whole genome amplification approach reveals novel polyhydroxyalkanoate synthases (PhaCs)
610   from Japan trench and nankai trough seawater. BMC Microbiol. 14, e11836.
611   doi:10.1186/s12866-014-0318-z
612   Fukui, T., Doi, Y., 1997. Cloning and analysis of the poly(3-hydroxybutyrate-co-3-
613   hydroxyhexanoate) biosynthesis genes of *Aeromonas caviae*. J. Bacteriol. 179, 4821–4830.
614   doi:10.1186/s12866-014-0318-z
615   Fukui, T., Shiomi, N., Doi, Y., 1998. Expression and characterization of (R)-specific enoyl
616   coenzyme A hydratase involved in polyhydroxyalkanoate biosynthesis by *Aeromonas caviae*. J.
617   Bacteriol. 180, 667–673.
618   Hoppensack, A., Rehm, B.H., Steinbüchel, A., 1999. Analysis of 4-phosphopantetheinylation of
619   polyhydroxybutyrate synthase from *Ralstonia eutropha*: generation of beta-alanine auxotrophic
620   Tn5 mutants and cloning of the panD gene region. J. Bacteriol. 181, 1429–1435.
621   Jia, Y., Kappock, T.J., Frick, T., Sinskey, A.J., Stubbe, J., 2000. Lipases provide a new mechanistic
622   model for polyhydroxybutyrate (PHB) synthases: characterization of the functional residues in
623   *Chromatium vinosum* PHB synthase. Biochemistry 39, 3927–3936. doi:10.1021/bi9928086
624   Kawashima, Y., Cheng, W., Mifune, J., Orita, I., Nakamura, S., Fukui, T., 2012. Characterization
625   and functional analyses of R-specific enoyl coenzyme A hydratases in polyhydroxyalkanoate-
626   producing *Ralstonia eutropha*. Appl. Environ. Microbiol. 78, 493–502.
627   doi:10.1128/AEM.06937-11
628   Keshavarz, T., Roy, I., 2010. Polyhydroxyalkanoates: bioplastics with a green agenda. Current
629   Opinion in Microbiology 13, 321–326. doi:10.1016/j.mib.2010.02.006
630   Matsumoto, K., Takase, K., Aoki, E., Doi, Y., Taguchi, S., 2005. Synergistic effects of Glu130Asp
631   substitution in the type II polyhydroxyalkanoate (PHA) synthase: enhancement of PHA
632   production and alteration of polymer molecular weight. Biomacromolecules 6, 99–104.
633   doi:10.1021/bm049650b
634   Matsusaki, H., Manji, S., Taguchi, K., Kato, M., Fukui, T., Doi, Y., 1998. Cloning and molecular
635   analysis of the Poly(3-hydroxybutyrate) and Poly(3-hydroxybutyrate-co-3-hydroxyalkanoate)
636   biosynthesis genes in *Pseudomonas* sp. strain 61-3. J. Bacteriol. 180, 6459–6467.
637   Meng, D.-C., Shen, R., Yao, H., Chen, J.-C., Wu, Q., Chen, G.-Q., 2014. Engineering the diversity
638   of polyesters. Curr. Opin. Biotechnol. 29, 24–33. doi:10.1016/j.copbio.2014.02.013
639   Noda, I., Green, P.R., Satkowski, M.M., Schechtman, L.A., 2005. Preparation and properties of a
640   novel class of polyhydroxyalkanoate copolymers. Biomacromolecules 6, 580–586.
641   doi:10.1021/bm049472m

642 Park, S.J., Kim, T.W., Kim, M.K., Lee, S.Y., Lim, S.-C., 2012. Advanced bacterial
643   polyhydroxyalkanoates: towards a versatile and sustainable platform for unnatural tailor-made
644   polyesters. Biotechnology Advances 30, 1196–1206. doi:10.1016/j.biotechadv.2011.11.007
645 Patil, K.R., Roune, L., McHardy, A.C., 2012. The PhyloPythiaS web server for taxonomic
646   assignment of metagenome sequences. PLoS ONE 7, e38581. doi:10.1371/journal.pone.0038581
647 Pärnänen, K., Karkman, A., Virta, M., Eronen-Rasimus, E., Kaartokallio, H., 2015. Discovery of
648   bacterial polyhydroxyalkanoate synthase (PhaC)-encoding genes from seasonal Baltic Sea ice
649   and cold estuarine waters. Extremophiles 19, 197–206. doi:10.1007/s00792-014-0699-9
650 Quandt, J., Hynes, M.F., 1993. Versatile suicide vectors which allow direct selection for gene
651   replacement in Gram-negative bacteria. Gene 127, 15–21. doi:10.1016/0378-1119(93)90611-6
652 Rehm, B.H.A., 2003. Polyester synthases: natural catalysts for plastics. Biochem. J. 376, 15–33.
653   doi:10.1042/BJ20031254
654 Sato, S., Kanazawa, H., Tsuge, T., 2011. Expression and characterization of (R)-specific enoyl
655   coenzyme A hydratases making a channeling route to polyhydroxyalkanoate biosynthesis in
656   *Pseudomonas putida*. Appl. Microbiol. Biotechnol. 90, 951–959. doi:10.1007/s00253-011-3150-
657   5
658 Schallmey, M., Ly, A., Wang, C., Meglei, G., Voget, S., Streit, W.R., Driscoll, B.T., Charles, T.C.,
659   2011. Harvesting of novel polyhydroxyalkanaote (PHA) synthase encoding genes from a soil
660   metagenome library using phenotypic screening. FEMS Microbiol Lett 321, 150–156.
661   doi:10.1111/j.1574-6968.2011.02324.x
662 Sheu, D.S., Wang, Y.T., Lee, C.Y., 2000. Rapid detection of polyhydroxyalkanoate-accumulating
663   bacteria isolated from the environment by colony PCR. Microbiology 146 (Pt 8), 2019–2025.
664   doi:10.1099/00221287-146-8-2019
665 Shozui, F., Sun, J., Song, Y., Yamada, M., Sakai, K., Matsumoto, K., Takase, K., Taguchi, S., 2010.
666   A new beneficial mutation in *Pseudomonas* sp. 61-3 polyhydroxyalkanoate (PHA) synthase for
667   enhanced cellular content of 3-hydroxybutyrate-based PHA explored using its enzyme homolog
668   as a mutation template. Biosci. Biotechnol. Biochem. 74, 1710–1712. doi:10.1271/bbb.100224
669 Simon, C., Daniel, R., 2011. Metagenomic analyses: past and future trends. Appl. Environ.
670   Microbiol. 77, 1153–1161. doi:10.1128/AEM.02345-10
671 Tai, Y.T., Foong, C.P., Najimudin, N., Sudesh, K., 2015. Discovery of a new polyhydroxyalkanoate
672   synthase from limestone soil through metagenomic approach. Journal of Bioscience and
673   Bioengineering. doi:10.1016/j.jbiosc.2015.08.008
674 Takase, K., Taguchi, S., Doi, Y., 2003. Enhanced synthesis of poly(3-hydroxybutyrate) in
675   recombinant *Escherichia coli* by means of error-prone PCR mutagenesis, saturation mutagenesis,
676   and in vitro recombination of the type II polyhydroxyalkanoate synthase gene. J. Biochem. 133,
677   139–145.
678 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular
679   Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
680   doi:10.1093/molbev/mst197
681 Tortajada, M., da Silva, L.F., Prieto, M.A., 2013. Second-generation functionalized medium-chain-
682   length polyhydroxyalkanoates: the gateway to high-value bioplastic applications. Int. Microbiol.
683   16, 1–15. doi:10.2436/20.1501.01.175
684 Tripathi, L., Wu, L.-P., Dechuan, M., Chen, J., Wu, Q., Chen, G.-Q., 2013. *Pseudomonas putida*
685   KT2442 as a platform for the biosynthesis of polyhydroxyalkanoates with adjustable monomer
686   contents and compositions. Bioresource Technology 142, 225–231.
687   doi:10.1016/j.biortech.2013.05.027

688 Verlinden, R.A.J., Hill, D.J., Kenward, M.A., Williams, C.D., Radecka, I., 2007. Bacterial synthesis
689   of biodegradable polyhydroxyalkanoates. J Appl Microbiol 102, 1437–1449.
690   doi:10.1111/j.1365-2672.2007.03335.x
691 Wu, H.-A., Sheu, D.-S., Lee, C.-Y., 2003. Rapid differentiation between short-chain-length and
692   medium-chain-length polyhydroxyalkanoate-accumulating bacteria with spectrofluorometry.
693   Journal of Microbiological Methods 53, 131–135. doi:10.1016/S0167-7012(02)00232-4
694 Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab initio gene identification in metagenomic
695   sequences. Nucleic Acids Research 38, e132–e132. doi:10.1093/nar/gkq275

696
697
698
699 **Figure legends**
700

701 Fig. 1. Activity-based screening of 11AW metagenomic library clones carrying PHA biosynthesis

702 genes in *P. putida* PpUW2 (PHA$^-$). PpUW2 recombinant strains were grown in 0.1N M63 medium

703 containing 0.5% Na octanoate and Nile red (0.5 μg/ml). Cosmid pJC8 and pJC67 (P*lac::phaC1$_{Pp}$*)

704 were used as negative and positive controls respectively. 11AW clones 14 and 16 were streaked on

705 the same selection plate to verify the PHA$^+$ phenotype initially screened on LB with 0.5% (w/v)

706 octanoate.

707

708 Fig. 2. Phylogenetic analysis of the polyhydroalkanoate synthases (PhaC) of 11AW metagenomic

709 library clones. Protein sequence alignments were performed using MUSCLE (Edgar, 2004).

710 Neighbour-jointing phylogenetic trees were generated with MEGA6 (Tamura et al., 2013). GenBank

711 accession numbers of 11AW PhaC proteins: PhaC2$_1$ (ALV86274), PhaC$_3$ (ALV86289), PhaC$_4$

712 (ALV86299), PhaC$_5$ (ALV86308), PhaC$_6$ (ALV86351), PhaC$_7$ (partial, ALV86358), PhaC$_8$

713 (ALV86364), PhaC$_{10}$ (partial, ALV86387), PhaC$_{14}$ (ALV86397), PhaC1$_{16}$ (ALV86417), PhaC2$_{16}$

714 (ALV86419), PhaC$_{18}$ (ALV86462), PhaC$_{19}$ (ALV86476), PhaC$_{20}$ (ALV86493), PhaC$_{21}$

715 (ALV86517), PhaC$_{23}$ (ALV86529), PhaC$_{25}$ (ALV86574), PhaC$_{27}$ (ALV86602), PhaC1$_{50}$

716   (ALV86626), PhaC2$_{50}$ (ALV86626), PhaC$_{51}$ (ALV86651), PhaC$_{P1N3}$ (ALV86715), PhaC$_{P2N8}$

717   (ALV86750), PhaC$_{P2N10}$ (partial, ALV86755), PhaC$_{P4N10}$ (ALV86768), PhaC$_{P8N7}$ (ALV86771).

718

719   Fig. 3. The *pha* genes and PhaC proteins in 11AW metagenomic DNA library clones. (A) The PHA^{+}

720   clones are classified into Class I and II based on the PhaC protein sequences. (B) Conserved amino

721   acids required for the activity of PHA synthases are indicated in closed triangles. Residues affecting

722   substrate specificity were marked by open triangles. The positions of amino acid residues were

723   numbered based on the sequence of PhaC1$_{16}$ [GenBank: ALV86417]. GenBank accession numbers

724   of PhaC proteins: PhaC$_{14}$, ALV86397; PhaC$_{18}$, ALV86462; PhaC$_{19}$, ALV86476; PhaC$_{25}$,

725   ALV86574; PhaC1$_{2}$, ALV86274; PhaC$_{20}$, ALV86493; PhaC1$_{16}$, ALV86417; *C. necator* PhbC1,

726   AAA21975; *P. putida* PhaC1, Q88D25; and *P.* sp. 61-3 PhaC1, BAA36200.

727

728   Fig. 4. Complementation of *S. meliloti* Rm11476 (*phbC*) with metagenomic clones. The cosmid

729   DNA of PHA^{+} clones encoding Class I and II PHA synthases was transferred to *S. meliloti* Rm11476

730   by conjugation. Recombinant strains were grown in YM medium. PHA production was estimated by

731   measuring the fluorescence of Nile red stained cells. *S. meliloti* wild type Rm1021 and previously

732   isolated Class I clones (CX92, pCX92M1, pCX92M3, and pCX92M5; (Schallmey et al., 2011))

733   were used as controls.

Figure 1.

Fig. 2

A

Class I

Clone 14 — $phaC_{14}$ 1716 bp, $phaA$ 1182 bp, $phaB$ 738 bp

Clone 18 — $phaC_{18}$ 1665 bp, $phaA$ 1182 bp, $phaB$ 738 bp

Clone 19 — $phaC_{19}$ 1692 bp, $phaA$ 1182 bp, $phaB$ 738 bp

Clone 25 — $phaC_{25}$ 1758 bp, $phaA$ 1179 bp, $phaB$ 738 bp

Class II

Clone 1 — $phaC1_1$ 834 bp, $phaZ$ 855 bp, $phaC2_1$ 1740 bp, $phaD$ 618 bp, $phaI$, $phaF$ 834 bp, $orf7$ 423 bp, 276 bp

Clone 16 — $phaJ_{16}$ 450 bp, $phaC1_{16}$ 1659 bp, $phaZ$ 957 bp, $phaC2_{16}$ 1140 bp

Clone 20 — $phaJ_{20}$ 468 bp, $phaC_{20}$ 1773 bp, $phaZ$ 837 bp

B

	120	229	287	318	391	445	473	478
	▽	▽	▼	▽	▼	▼	▼	▽

Class I
$PhbC1_{re}$ ...SQWVDAMSP......QPESSLVRH......INVLGFCVGGTIVS......TLLTTLLDF...NDLVWNYVV......GSREDHIVP......GASGHIAGVINP......PQQ...

$PhaC_{25}$ ...EQWMAASSP......QPENSLIRY......VNVLGFCVGGTILS......TLLTTFLDF......NDLVWNYVV......GSREDHIVP......GASGHIAGVINP......QAE...

$PhaC_{14}$ ...QQWVDAMSP......QPENSLIRY......FNMLGFCVGGTLIA......TLLTTLLDF......NDLVWNYVV......GSREDHIVP......GASGHIAGVINP......ADE...

$PhaC_{18}$ ...QQWIDAASP......QPDNSLIRW......IDVLGFCVGGTILA......TLLTTLLDF......NDLVWNYVV......ASREDHIVP......GASGHIAGVINP......AQQ...

$PhaC_{19}$ ...QQWVDAASP......QPDNSLIRY......IDTLGFCVGGTILT......TLLTTLLDF......NDLVWNYVV......GSREDHIVP......GASGHIAGVINP......AQQ...

Class II
$PhaC1_{16}$ ...TMLVEALAP......SPGNSFVNY......VSFKGLCAGGLTSA......TLNVTLLDI......NDLVWNYWV......AGRTDHITP......VNSGHIQTLVCP......PHE...

$PhaC2_1$ ...ALLNDAMAP......SPHNSFVQF......VNLMGACAGGLTIA......TYLVSLLDS......NDLIWSYFV......AGINDHITP......SNSGHVQSILNP......PRA...

$PhaC_{20}$ ...SLMTEALAP......SPGKSLVEY......VNTLAACAGGITLM......TLLVTLLDT......NDLVWNYWV......AGTTDHITP......SSSGHIQSIVNP......PDD...

$PhaC1_{Pp}$ ...TLMTEAMAP......SPHNSFVQF......VNLMGACAGGLTIA......TYLVSLLDS......NDLIWSYFV......AGINDHITP......SNSGHVQSILNP......ANE...

$PhaC1_{Psp.\ 61-3}$ ...NLMTEAMAP......SPDKSLARF......INMLGACSGGITCT......TLLVSVLDT......NDLIWNYWV......AGTNDHITP......SSSGHIQSILNP......PVA...

Lipase box

Fig. 3

33

Fig. 4

Table 1. Bacterial strains, plasmids and cosmids

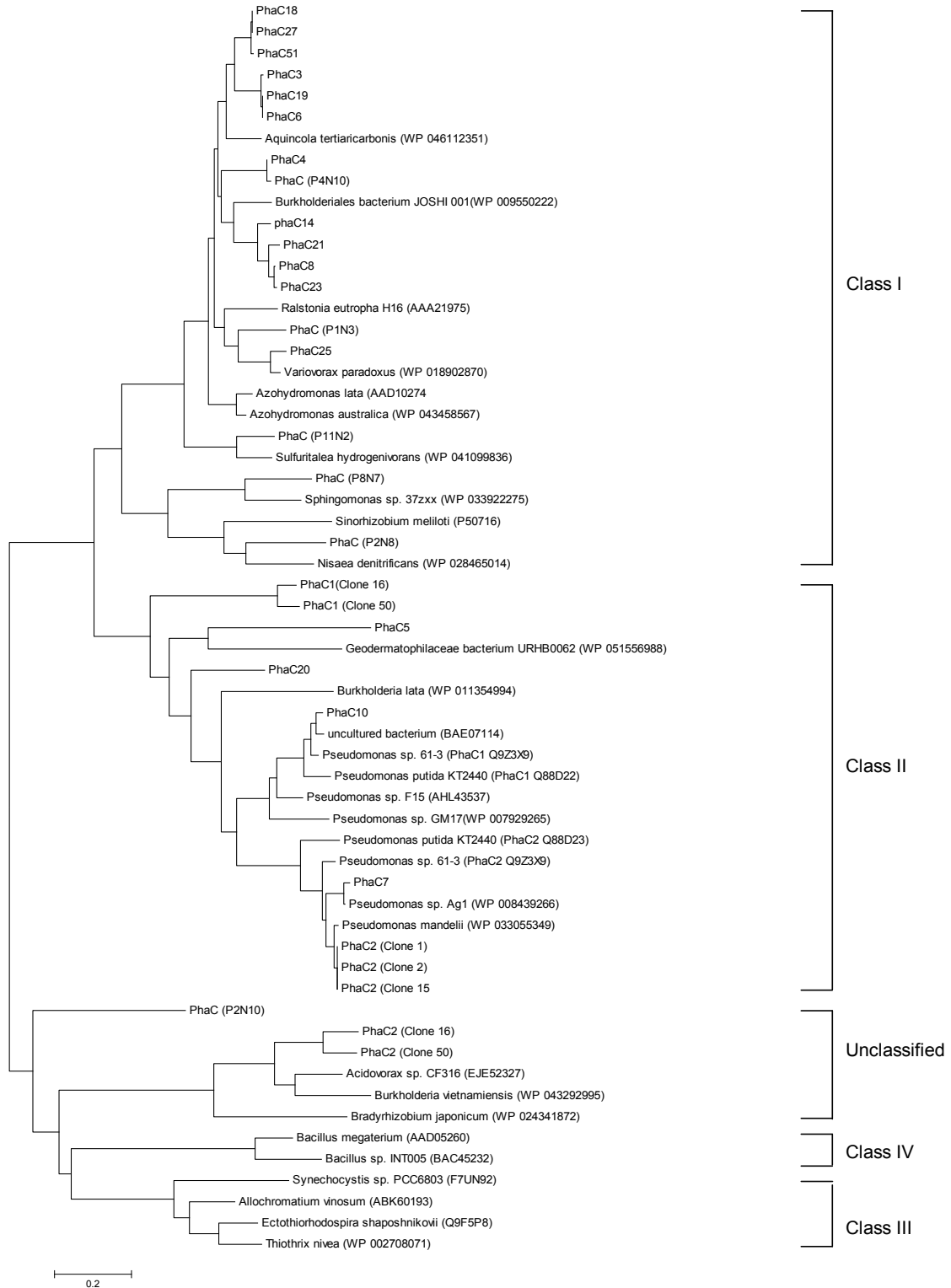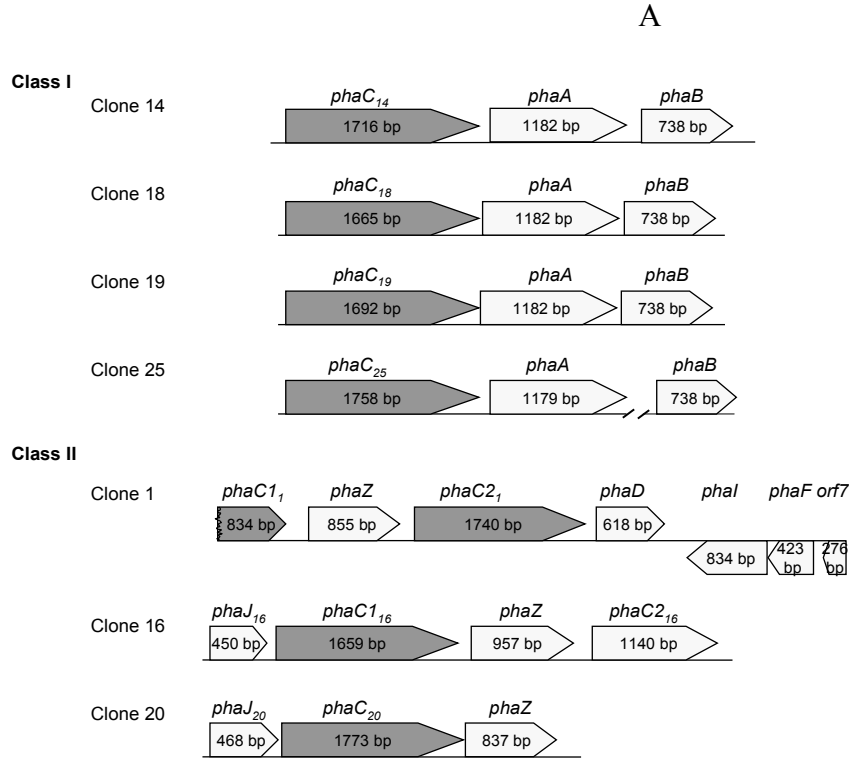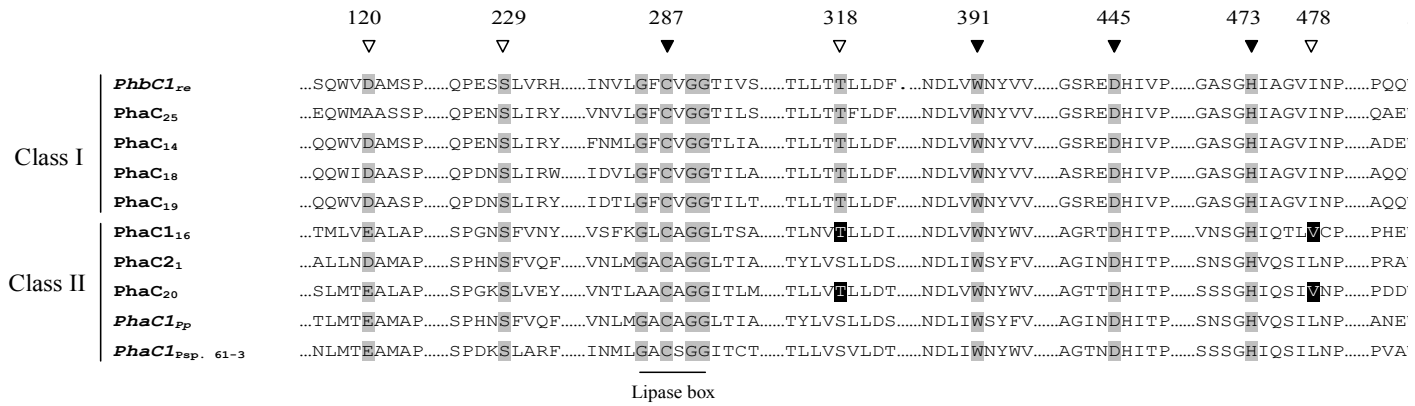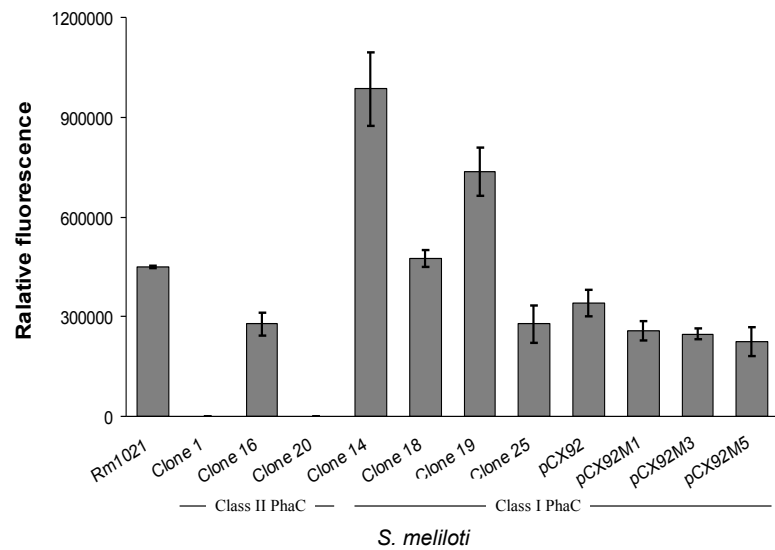| Strain, plasmids and cosmids	Relevant characteristics*	Reference
E. coli		
DH5α	F⁻ φ80 *lacZΔM15Δ*(*lacZYA-argF*)*U169 recA1 endA1 hsdR17 phoA supE44 thi-1 gyrA96 relA1*	Life Technologies
DH5α (Rif^R)	A spontaneous Rif^r mutant of DH5α	(Cheng et al., 2014)
HB101	F⁻ *thi-1 hsdS20 supE44 recA13 ara-14 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	CGCCAAGCTTG**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	CGCGAAGCTTG**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* (â-Proteobacteria)	Hypothetical protein (WP_046112351), *Aquincola tertiaricarbonis*	74	KT944255
4	40,585	65.7	*Leptothrix* (â-Proteobacteria)	Polyhydroxyalkanoate synthase (AAD10274), *Azohydromonas lata*	72	KT944256
6	>11,479	67.7	*Leptothrix* (â-Proteobacteria)	Polyhydroxyalkanoate synthase (AAD10274), *Azohydromonas lata*	73	KT944258
8	34,220	67.8	*Leptothrix* (â-Proteobacteria)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	74	KT944260
14	29,565	69.1	*Methylibium* (â-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* (â-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* (â-Proteobacteria)	Poly-(R)-hydroxyalkanoic acid synthase (WP_009550222), Burkholderiales bacterium JOSHI_001	74	KT944268
25	41,012	66.8	*Varivorax* (â-Proteobacteria)	Poly-(R)-hydroxyalkanoic acid synthase (WP_018902870), *Variovorax paradoxus*	93	KT944269
27	37,156	67.0	*Rubrivivax* (â-Proteobacteria)	Hypothetical protein (WP_043458567), *Azohydromonas austrilica*	73	KT944270
51	38,794	69.9	*Rubrivirax* (â-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* (â-Proteobacteria)	Poly(3-hydroxyalkanonate) synthase (WP_041099836), *Sulfuritalea hydrogenivorans*	73	KT944278
P1N3	32,797	68.6	*Variovorax* (â-Proteobacteria)	Poly(3-hydroxyalkanonate) synthase (WP_028253507), *Variovorax* sp. URB0020	86	KT944273
P4N10	>2,724	64.2	*Leptothrix* (â-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
Class II						
16 (PhaC1₁₆)	33,900	66.1	*Methylibium* (â-Proteobacteria)	PhaC1 synthase (AHL43537), *Pseudomonas* sp.F15	48	KT944263
50 (PhaC1₅₀)	22,965	67.0	*Hydrocarboniphaga* (ã-Proteobacteria)	Polyhydroxyalkanoic acid synthase (WP_007929265), *Pseudomonas* sp. GM17	46	KT944271
1 (PhaC1₂)	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* (â-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* (â-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 (PhaC2₁₆)	33,900	66.1	*Methylibium* (â-Proteobacteria)	Easterase (WP_043292995), *Burkholderia vietnamiensis*	53	KT944263
50 (PhaC2₅₀)	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.

37

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).

Supplementarys

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
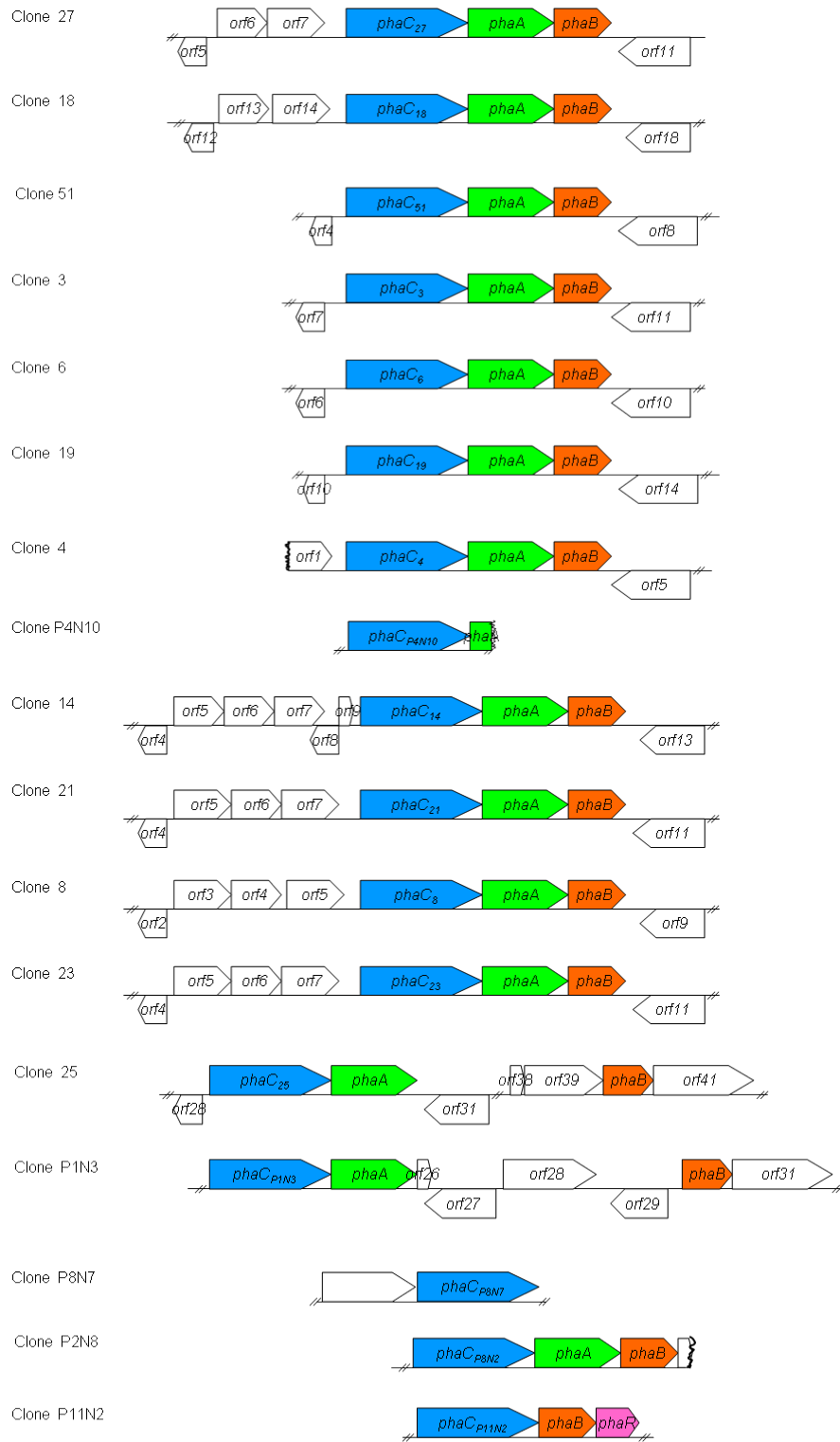
Class II
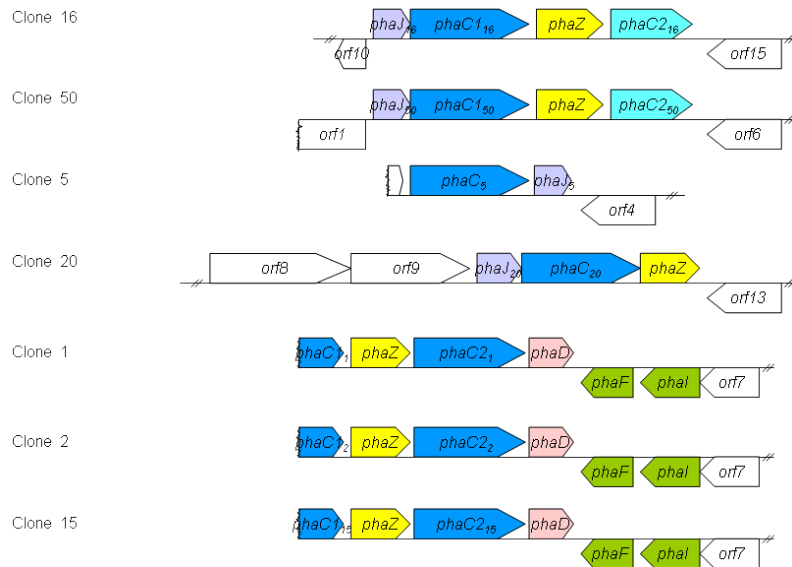
Unclassifed
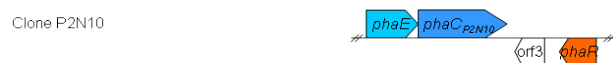
Supplementary 1.
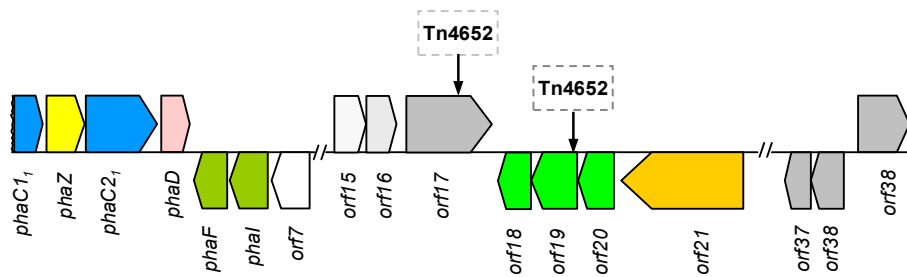
Supplementary 2.

```
PhaC2₁₆       -----msslpvaers--------------------------------------------
PhaC2₅₀       ----Mistkpvaers--------------------------------------------
PhaC_P2N10*   ----------------------------------------------------------
PhaC1₁₆       MAktkgsrrAriPsP----psrS----------EmLrlAL------------------
PhaC1₅₀       ----MaKtkApprwarlpsppsrs----------ElLrlALv-----------------
PhaC5         MAkreatknvAatev------------------vggGedVG------------------
PhaC20        --mtatRrskArPde-prakkrv----------rkVaaAVGapkpeagEArgfae----
PhaC10*       ----------------------------------------------------------
PhaC2₁        --mrektarefsPtPaafinaqS----------aitGlrgr-----------------
PhaC7*        ----------------------------------------------------------
PhaC_P11N2    ----MqyaghAlfqs------fn----------QlLaaqLqQ----------------
PhaC25        ----MmKqqAtgada---fapfq----------QALsegwnKAlesfqqSatqgasafn
PhaC_P1N3     ----MrQeqqAaada--tfapfq----------QALtqgwe-----------------
PhaC14        ----MkskpAAsarPtkqvtDsA----------QALGhAwG-----------------
PhaC8         ----MkskpAApsrP-dqvaEsA----------QALGhAwG-----------------
PhaC23        ----MkskpAApPrP-dqvaEsA----------QALGhAwG-----------------
PhaC21        ----MkskpAApPrP-dqvaEsA----------QsLGhAwG-----------------
PhaC4         ----MpKtiASlPas-------A----------QAyGaAVG-----------------
PhaC_P4N10    ----MpKtiASlPas-------A----------QAyGaAVG-----------------
PhaC19        ----MsKnntklPhs-----EqA----------dAfGhAVG-----------------
PhaC6         ----MsKnntklPhs-----EqA----------dAfGhAVG-----------------
PhaC3         ----MsKnntklPhs-----EqA----------dAfGhAVG-----------------
PhaC51        ----MmtgdA----------------------dAfGkAMG-----------------
PhaC27        -----mtgdA----------------------dAfGKAMG-----------------
PhaC18        -----mtgdA----------------------dAfGkAMG-----------------
PhaC_P8N7     MAdpdtKapeApPlP---slEem----------QhwtwvMGRAqqmmlE---------
PhaC_P2N8     ----MqtenSApidPgavleElAqvaqqiqgqmrdALdggwqQpvvdg----------

PhaC2₁₆       -----nLeKlL--------LAvP------------------------------ayW---
PhaC2₅₀       -----nLeKAL--------LAfp------------------------------pyW---
PhaC_P2N10*   ------------------MiLP---------glAaAaL------QAE--------rer
PhaC1₁₆       -----vilgrLrGn-----pgLP-------------------------IRhiagfakef
PhaC1₅₀       -----ifgRlrrn------pALP-------------------------IRhiaafakel
PhaC5         ------vlsprSl------Ieaa------------tsAMNpL--------tmaRESarLygew
PhaC20        -----rvvdAIpG-------pnPfvgfSvFdvlssaeQLvagalqQpglVlrnqaafaGd
PhaC10*       ----------------------------------------------------------
PhaC2₁        -----dLFsTLrs------------------vaAhglr------npvhtarhalklgGq
PhaC7*        ----------------------------------------------------------
PhaC_P11N2    -----ssveltnG------LSht------gtaasasvmaL------QeEwqarhmQLWqGm
PhaC25        vtggtpLWqipqt------aSMP-elpkisidpekLqsi------QqQYVaEATELWrqg
PhaC_P1N3     -----kaleAfqslgKApqadMPwdmprfsfsasrLqEL------QrhYIeEATELWrqd
PhaC14        -----eMWKSavG------LSLP----------tdAMSEL------QtrYVKQATELWNqa
PhaC8         -----hMWKSVAG------MSLP----------atAMTEL------QSnYLKQATELWNqt
PhaC23        -----dMWKSVAG------MSLP----------atAMTEL------QSnYLKQATELWNqt
PhaC21        -----dMWKSVAG------MSLP----------atAMTEL------QSnYLKQATELWNqt
PhaC4         -----diWKSMSe------LNLP---------lPALSEL------QSnYVKQATEIWNGa
PhaC_P4N10    -----diWKSMSe------LNLP---------lPALSEL------QSnYVKQATEIWNGa
PhaC19        -----diWKSMqG------LSLP---------lPALSEL------QtEYLKDATaLWNGa
PhaC6         -----diWKSMqG------LSLP---------lPALSEL------QtEYLKDvTaLWNGa
PhaC3         -----diWKSMqG------LSLP---------lPALSEL------QtEYLKDATaLWNGa
PhaC51        -----rMlqAaAG------LTLP---------ptALTDL------QAEYLREATEMWNGa
PhaC27        -----rMlqAaAG------LTLP---------psALTDL------QAEYLREATEMWNGa
PhaC18        -----rMlqAaAG------LTLP---------psALTDL------QAEYLREATEMWNGa
PhaC_P8N7     -----hvaRAMgetaEAkapALP--wlTlFgdpakLAQq------QAElwtEglaIWqra
PhaC_P2N8     -----pvlnSMAG---gllqALavfnqSlFadparLlQa------QAKlwdgyaQLWqrg

PhaC2₁₆       ----------------------piEW----------------------------ar
PhaC2₅₀       -----------------------piEW---------------------------tk
PhaC_P2N10*   LDaLlgasaea-------------vsWpdrsAAcLpv------------------eg
PhaC1₁₆       LniLrgksdv----tpaagDkRFAdaaWSgNrlSgalvQvYiaasqeayaLvkeLEl-eg
PhaC1₅₀       LrvLrnksev----ApsagDkRFtdaaWSdrrwSSalvQvYLaasqeayaLvneLEl-Dg
PhaC5         LKiLvgrser----dvPvkDwRFAdptWhehPAykrTAQgYLafcdaidRviDgnp--Dw
PhaC20        LaRIlsgrstl---epeegDRRFqdptWkeNPfyraglQtYLawrkgvhRLvegagl-Da
PhaC10*       -------------------------------------------------------
PhaC2₁        LgRvllgetlh---ptnpqDnRFAdpaWSlNPfyrrSlQaYLswqkevkhwiDdsnm-tp
PhaC7*        -------------------------------------------------------
PhaC_P11N2    LQsnqaatPva---qslpgDRRFshpaWAesPiydylrQaYLINAdyLqRLADtap--eg
PhaC25        fa-----------AkPegDRRFASdaWgSNPlSAfSAavYLLNgRTLLnMADAIDa-De
PhaC_P1N3     ia-----------ArthgDRRFAAeaWgSNPvAAfSAavYLLNtRTLLgMAeAVEG-Da
PhaC14        LaRLqpDGakETAPAaPitDkRFAApEWSANPATAyTAQmYLLNaRTLMQLAesMQG-Da
PhaC8         LsRLqpEGatETAaAtPitDkRFAApEWSANPASAyTAQmYLLNaRTLMQLAesMQG-Da
PhaC23        LsRLqpEGatETAaAtPitDkRFAApEWSANPASAyTAQmYLLNaRTLMQLAesMQG-Da
PhaC21        LsRLqpEGatETAPAtPitDkRFAApEWSAsPASAyTAQmYLLNaRTLMQLAesMQG-Da
PhaC4         vERLqaDGets---kkPlpDRRFAAnEWAANPvSALTAQiYLLNaRTLLQMADAIEG-De
PhaC_P4N10    vERLqaDGets---kkPlpDRRFAAnEWAANPvSALTAQiYLLNaRTLLQMADAIEG-De
PhaC19        vERLktDesqt---AvaprDRRFAAdEWASNPAAAMaQtYLLNaRTLMKMADAVEG-Dt
PhaC6         vERLktDeskt---AvaprDRRFAAdEWASNPAAAMaQtYLLNaRTLMKMADAVEG-Dt
PhaC3         vERLktDeskt---AvaprDRRFAAdEWASNPAAAMaQtYLLNaRTLMKMADAVEG-Dt
PhaC51        vERfsgkaPeQ---AqkigDRRFAAnDWAANPATALaAQaYLLNsRTLMKMADAIQG-Da
PhaC27        vERfsgkaPeQ---ShkigDRRFAAnDWAANPATALaAQaYLLNsRTLMKMADAIQG-Da
PhaC18        vERfsgkaPeQ---ShkigDRRFAAnDWAANPATALaAQaYLLNsRTLMKMADAIQG-Da
```

```
PhaC_P8N7          Lg--vdsGPrtelvekadrDRRFAApQWreNPlfdMirQsYLLvserLLgsvDsIDGLDp
PhaC_P2N8          LgaLtgqvPpaagPptaptDRRFrAqDWql-PAfdLirQwYLLtsqwLLEqvtgagdLDp

                                        120
PhaC2₁₆            qiRkRgtklprdWlryVraagaItf-PD----------------------Rpd--wVt
PhaC2₅₀            qiRqRgmrlpldWVsyVreakaItf-PD----------------------KpG--wVt
PhaC_P2N10*        KaR---------------------------------------------------------
PhaC1₁₆            rdaSaaRFlstmlVeALaPSNnpltNPvvlKrvrETRGkSllrGfQHfLsDrldnaGmIS
PhaC1₅₀            rdaSsaRFlatmlVeAMaPSNnpltNPvvmqrvrETKGrSllrGLQHfLsDrldnaGmIS
PhaC5             rkReRaRFltgiltsAMaPtNtLigNPaAlKKAyETgGkSlvrGvQnLIDDVRhnkGmpS
PhaC20            KgteRaRFvaslmteALaPtNFflgNPaAiKrALETgGgSllrGLanLfDDVatngGmpS
PhaC10*          ---------------------------------------------------------
PhaC2₁            ddRARahFAfallnDAMaPSNsL-LNPlAiKeifnsgGnSlvrGisHLvDDllhndGlpr
PhaC7*           ---------------------------------------------------------
PhaC_P11N2        qiRnRMqFlthQaIDAMaPSNFaAtNPEfiKtALaTeGESIrlGiQnLLgDlEkG--rIS
PhaC25            KTkARLRFAVEQWmaAsSPSNsLAfNaEAQKKAIdTqGESIAkGiQnLLhDVKQG--HlS
PhaC_P1N3         KTRARLRFAVEQWVaAsSPSNYLAfNaEAQKKALEsqGESIAkGiQnLLhDVQQG--HlS
PhaC14            KTkARVRFAVQQWVDAMSPaNFMALNPEAQKKALdTKGESVSQGiQqLLNDlRkG--aVS
PhaC8            KTkARVRFAVQQWVDAMSPaNFMALNPEAQKKALdTKGESVAQGLQqLfNDlRkG--sVS
PhaC23           KTkARVRFAVQQWVDAMSPaNFMALNPEAQKKALdTKGESVAQGLQqLfNDlRkG--sVS
PhaC21           KTkARVRFAVQQWVDAMSPaNFMALNPEAQKKALdTKGESVAQGLQqLfNDlRkG--sVS
PhaC4            KTRqRIRFAVQQWVeAaSPSNYLALNPEAQrKALETKGESItQGLlHLwqDVQQG--HVS
PhaC_P4N10        KTRqRIRFAeQQWgeAarrSNYvALNPEAQrKALETKGESItQGLlHLwqDVQQG--HVS
PhaC19           KTRARIRFAVQQWVDAaSPSNYLALNPEAQrKALETRGESIAQGLRqLwDDIQQG--HVS
PhaC6            KTRARIRFAVQQWVDAaSPSNYLALNPEAQrKALETRGESIAQGLRqLwDDIQQG--HVS
PhaC3            KTRARIRFAVQQWVDAaSPSNYLALNPEAQrKALETRGESIAQGLRqLwDDIQQG--HVS
PhaC51           KTkARVRFAVQQWIDAaSPSNYLAfNPEAQrKALETKGESIAQGLaHLLgDIQQG--HVS
PhaC27           KTkARVRFAVQQWIDAaSPSNYLAfNPEAQrKALETKGESIAQGLaHLLgDIQQG--HVS
PhaC18           KTkARVRFAVQQWIDAaSPSNYLAfNPEAQrKALETKGESIAQGLaHLLgDIQQG--HVS
PhaC_P8N7         KqRekLRFmtrnfVDAMaPSNFaAtNPvvlerAMaTRGESllkGLEHmLDDVskG--qlt
PhaC_P2N8         aTRhkLaFytrQfaDALSPSNFvAtNPEvlratvEsgGqnllhGfRHvLEDmKrGgGrId

PhaC2₁₆           ------------------pntiayEsqtmlLrnfsdpvvsadtiPviidaPfaghPstIa
PhaC2₅₀           ------------------phtiaFEtqtmwLrdfsdadApadaiPvivdaPyaghPstIa
PhaC_P2N10*       ----------------------------LLcYpsLrepa-aRPLLlVyslvNr-ihvL
PhaC1₁₆           --mvDrtaFEVGRNVAaTpGdVVFrNEvleLLqYKPqsAeVyEiPLmvVPPqINK-YYvL
PhaC1₅₀           --mvDrtpFEVGRNVAcTpGdVVFrNEvcevLqYKPqTAeVysiPLvvVPPqINK-fYvL
PhaC5            --QvkrSdFkVGeNlAaTpGAVVlrtErleLLqYqPtTAKVyaiPtLiIPPpIgK-hYfi
PhaC20           --QvDkSaFrVGeNlAsTpGAVVFrNEvleviqYaPagAqVygRPLLlVPPqINK-YYvi
PhaC10*          ---------------------------------------------------------
PhaC2₁           --QvtkqaFEVGktVATTtGsVVFrNELleLiqYKPMseKqysKPLLvVPPqINK-YYIf
PhaC7*           ---------------------------------------------------------
PhaC_P11N2        --mTDEaaFEiGRNlAvtTsGsVVFENELmQLiqYaPLTpKVaaRPLviVPPCINK-fYIL
PhaC25           --mTDESaFEVGRNVATTEGAVVFENELFQLLEYKPLTpKVyERPfLlVPPCINK-YYIL
PhaC_P1N3        --mTDESaFEVGRNVATTEGAVVFENELFQLLEYKPLTAKVyERPLLlVPPCINK-fYIL
PhaC14           --QTDEtVFEVGRNVATsEGAVVFENELFQLiEYKPLTgKIHtRPMLFVPPCINK-YYIL
PhaC8           --QTDESVFEVGRNVATsEGAVVFENELFQLiEYKPLTAKVHtRPMLFVPPCINK-fYIL
PhaC23          --QTDESVFEVGRNVATsEGAVVFENELFQLiEYKPLTAKVHtRPMLFVPPCINK-fYIL
PhaC21          --QTDESVFEVGRNVATsvGAVVFENELFQLiEYKPLTAKVHtRPMLFVPPCINK-fYIL
PhaC4           --QTDESVFEVGRNVATTEGAVVFENdLFQLiEYKPLTdKVHEtPfLFVPPCINK-YYIL
PhaC_P4N10       --QTDESVFEVGRNVATTEGAVVFENdLFQLiEYKPLTdKVHEtPfLFVPPCINK-YYIL
PhaC19          --QTDESVFEVGRNVATsEGAVVFENELFQLLEYKPLTAqVHERPMLFIPPCINK-YYIL
PhaC6           --QTDESVFEVGRNVATsEGAVVFENELFQLLEYKPLTAqVHERPMLFIPPCINK-YYIL
PhaC3           --QTDESVFEVGRNVATsEGAVVFENELFQLLEYKPLTAqVHERPMLFIPPCINK-YYIL
PhaC51          --QTDESVFEVGRNVATTEGAVVFENELFQLLEYKPLTAqVHERPMLFVPPCINK-YYIL
PhaC27          --QTDESVFEVGRNVATTEGAVVFENELFQLLEYKPLTAqVHERPMLFVPPCINK-YYIL
PhaC18          --QTDESVFEVGRNVATTEGAVVFENELFQLLEYKPLTAqVHERPMLFVPPCINK-YYIL
PhaC_P8N7        --hTDpeaFEVGRNiAvTpGkVVkrtpLYeLiqYsPtTetVygtPLvifPPwINr-fYIL
PhaC_P2N8        prmTDpnaFEVGaNiAaTpGkVVFqNgLmQLiqYtPvTdqVyrRPLLvVPPwINK-fYvm

                                        229
PhaC2₁₆           DyadgqSLvqtlkagGCkRVyVtdWksatheMrnfgiDkYlsDld----QaIdAVgGRa-
PhaC2₅₀           DyadgqSLvqtlkagGCkRlyVtdWksathaMrnfgiDkYlaD--LR--daVdhVgGRa-
PhaC_P2N10*       DLarghSfIerlqglG-HpVyVldWgrPDaSdAglslaDYVcgylstAVRLVrrhgcaR-
PhaC1₁₆           DLsPgNSfvnYaVKQG-fqVFmVSWRNPgpeLAhlslDhYlr-GleeAhnaVmeITGaK-
PhaC1₅₀           DLsPgNSfvnYaVKQG-fqVFmVSWRNPgpeLAylslqhYlQ-aledAhRaVmeITGad-
PhaC5            DLsPgrSfveYaVakG-fqVFcISWRNPqveqgdwgiDDYI-gsmLeAVdaVckITGsK-
PhaC20           DLsPgkSLveYaVasG-iqfFaVSWRNPtvaerdwgletYll-alseAIdaVceITGsE-
PhaC10*         -------------------FmVSWRNPtkaqrewglstYIE-alkeAVdVVsAITGsK-
PhaC2₁          DLsPhNSfvqfalKnG-lqtFmISWRNPDvrhrewglssYVE-aveeAmnVcrAITGaR-
PhaC7*          -------------------VFVVSWRNPDvrhrewglstYVa-aleeAlnVtrAITGaR-
PhaC_P11N2       DLQPENSLvRYaVEQG-ntVFmVSWRNvqadLghlTWDDYIEqGaLKAIRVaQqIcrvp-
PhaC25          DLQPENSLIRYanEQG-HRVFVVSWRNPDeSLAnaTWDDYIEnaaIKAIhtVQdISGsK-
PhaC_P1N3       DLQPENSLIRYaVEQG-HRtFVVSWRNPDaSMsqyTWDDYIEDaaIKAIQVtreITGQdk
PhaC14          DLQPENSLIRYTVEQG-HRVFVmSWRNPDeSLAayTWDDYIEkGpIKAIeVVQqISGQK-
PhaC8           DLQPENSLIRYTVEQG-HRVFVmSWRNPDeSLAhfTWDDYIEkGpLKAIeVVQqIgGQK-
PhaC23          DLQPENSLIRYTVEQG-HRVFVmSWRNPDeSLAhfTWDDYIEkGpLKAIeVVQqIgGQK-
PhaC21          DLQPENSLIRYTVEQG-HRVFVmSWRNPDeSLAhfTWDDYIEkGpLKAIeMVQqIgGQK-
PhaC4           DLQPENSLIRwTVEQG-HRlFVmSWRNPDdSVAakTWnDYIEDGtIRAIeVreITGaE-
PhaC_P4N10       DLQPENSLIRwTVEQG-HRlFemSWRNPDdSVAakTWnDYIEDGtIRAIeVreITGaE-
PhaC19          DLQPDNSLIRYTVaQG-HRtFVVSWRNPDpSIAakTWDDYIDDGaIRAIReVQAISGQa-
```

```
PhaC6        DLQPDNSLIRYTVaQG-HRtFVVSWRNPDpSIAakTWDDYIDDGaIRAIReVQAISGQa-
PhaC3        DLQPDNSLIRYTVaQG-HRtFVVSWRNPDpSIAakTWDDYIDDGaIRAIReVQAISGQa-
PhaC51       DLQPDNSLIRwTVgQG-HRtFVVSWRNPDaSIAdkTWDDYIEDGaIRAIgeVQSISGQK-
PhaC27       DLQPDNSLIRwTVgQG-HRtFVVSWRNPDaSIAdkTWDDYIEDGaIRAIReVQSISGQK-
PhaC18       DLQPDNSLIRwTVgQG-HRtFVVSWRNPDaSIAdkTWDDYIEDGaIRAIReVQSISGQK-
PhaC_P8N7    DLnPKkSfIRwaVEQG-ltVFVVSWksaDeSLAgtTlDDYVlgGqVdAIdtIrellGvE-
PhaC_P2N8    DLQPKNSLvgwwlQQG-ytVFmISWiNPgpqLAeksfeDYmiEGpLaAldaIEqacGER-
```

```
                           287                       318
                       Lipase box
PhaC2₁₆      -----hlvGlCqGGwLs----AMhAARfp-gkAvSLvLagapiDthAgh-Gpvtkmaqrt
PhaC2₅₀      -----hlvGlCqGGwLst----MLAARyp-dkvvSLvLagapiDthAgD-Gpvtklskrt
PhaC_P2N10*  ---tIdlLGiCqGGs-LAlcLAsLepRal-----nrLvtLvTpvDFh--tpG--DVlaDlA
PhaC1₁₆      ---KVsfkGlCaGGltsAiALgryAdqGkLdtvntLTLnvTLLDiSgmEktnmgyFLtpe
PhaC1₅₀      ---KVNfkGlCaGGltLAiALgryAdqGkLdtvntLTLnvTLLDiAgmEktnmgyFLtpS
PhaC5        ---KVNisGFCaGGimttlmLsymAAvrD-krvnaaafgvmLLDfd--teapigalhsrk
PhaC20       ---RVNTLaaCaGGitLmllLsyLAAeGD-rrvnSvTLLvTLLDte--aeGvLgLFasEe
PhaC10*      ---dINmLGaCsGGitctalLghyAAlGE-kkvnaLTLLvsvLDtT--ldtqvaLFVDEq
PhaC2₁       ---EVNlMGaCaGGltIAalqghLqAkrQLrrvSSaTyLvsLLDsq--mdspatLFaDEq
PhaC7*       ---EVNlMGaCaGGltIAalqghLqAkrQLrriSSasyLvsLLDsq--idspatLFaDEq
PhaC_P11N2   ---QVNaLGFCVGGTILssALAVarlRGE-dPvASLTLLTTLLDFS--DsGeigyFVDEA
PhaC25       ---QVNvLGFCVGGTILsTALAVLAARGE-kPASSvTLLTTfLDFS--DTGiLDIFVDEp
PhaC_P1N3    -kggQIdTLFCVGGTILsTALAVLAARGE-QPASSvTLLTTLLDFe--DTGiLDIFIDEg
PhaC14       ---tfNmLGFCVGGTLIATALAVLAARGQ-QPAhSLTLLTTLLDFk--hTGvLDVFVDEp
PhaC8        ---QfNmLGFCVGGTLIATALAVLAARGE-QPAhSLTLLTTLLDFk--hTGvLDVFVDEp
PhaC23       ---QfNmLGFCVGGTLIATALAVLAARGE-QPAhSLTLLTTLLDFk--hTGvLDVFVDEp
PhaC21       ---EfNmLGFCVGGTLIsTALAVLAARGE-QPAhSLTLLTTLLDFk--hTGvLDVFVDEp
PhaC4        ---RlNTLGFCVGGTILATALAVLAARGD-QPAASvTLLTaLLDFS--ETGvLDLFIDEp
PhaC_P4N10   ---RlNTLGFCVGGTILATALAVLAARGD-QPAASvTLLTaLLDFS--ETGvLDLFIDEp
PhaC19       ---tIdTLGFCVGGTILtTALAVLAARGE-QPAASLTLLTTLLDFA--DTGvLDIFIDEA
PhaC6        ---tIdTLGFCVGGTILtTALAVLAARGE-QPAASLTLLTTLLDFA--DTGvLDIFIDEA
PhaC3        ---tIdTLGFCVGGTILtTALAVLAARGE-QPAASLTLLTTLLDFg--DTGvLDIFIDEA
PhaC51       ---QIdvLGFCVGGTILATALAVLAARGE-QPAASLTLLTTLLDFS--NTGvLDIFIDEA
PhaC27       ---QIdvLGFCVGGTILATALAVLAARGE-QPAASLTLLTTLLDFS--NTGvLDIFIDEA
PhaC18       ---QIdvLGFCVGGTILATALAVLAARGE-QPAASLTLLTTLLDFS--NTGvLDIFIDEA
PhaC_P8N7    ---sVhaIGYCVaGTtLAatLALLeARGEadkvASaTffTaqvDFS--EaGdLnLFVade
PhaC_P2N8    ---EVNavGYCIGGTLLllstLAymtAvnD-ErikSaTtfasLLDFS--DpGdLGVFIDEA
```

```
PhaC2₁₆      pfnfyrslVa----mGgGrmRGrfmlqgwknmhPdehyWgkyidlfdnitnPqyveRARH
PhaC2₅₀      plsfyrsmVa----mGgGrmRGrfmlqgwknmhPdehyWgkyidlfdnitnPqyvsRARH
PhaC_P2N10*  rVldlpalLapd-----GnLpGavLAglFgmLRPlrgVgaarapwldgGaEPekleRlRr
PhaC1₁₆      glEtsmkrskKE-----GvLyGhEmAkmFawLRPNDLVWNYwVnNYLmGQKPaaFD----
PhaC1₅₀      glEksmkrskQE-----GvLyGhEmAkmFawLRPNDLVWNYwVnNYLmGKKPaaFD----
PhaC5        lIgVarsrsaKk-----GIhpasaLAqvFawmRPNDLVWNYwVnNYLtGQdPPsFD----
PhaC20       aVaLaklnswtk-----GvLpGEELgrvFawLRPNDLVWNYwVnNYLlGnaPPaFD----
PhaC10*      tlEaakrhsyQa-----GvLEGrDmAkvFawmRPNDLi----------------------
PhaC2₁       tlEaakrrsyQk-----GvLdGrDmAkvFawmRPNDLiWsYfVnNYLlGKEPPaFD----
PhaC7*       tlEaakrrsyQQ-----GvLdGrDmAkvFawmRPNDLi----------------------
PhaC_P11N2   sVaaREagIG----QG-GLLhGrELATvFSaLRaNDLiWqYVVGNYLKGgKPPaFD----
PhaC25       mVayREmqLG----kG-GLLpGmDLAsTFSFLRPNDLVWNYVVGNYLKGEtPPPFD----
PhaC_P1N3    fVrfREmqMn----aG-GLLpGsELAsTFSFLRPNDLVWNYVVGNYLKGEtPPPFD----
PhaC14       sVQLREvTIGEQAPnGPGLLKGkELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC8        aVQMREmTLGEQAPnGPaLLKGkELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC23       tVQMREmTLGEQAPnGPaLLKGkELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC21       sVQMREmTIGEQAPnGPaLLKGkELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC4        aVQMREaTIGpQsPnGcGLLKGQELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC_P4N10   aVQMREaTIGpQsPnGcGLLKGQELATTFSFLRPNDLVWNYVVGNYLKGEaPPPFD----
PhaC19       aVQLREmTLGEQAPQGPaLLKGQELATTFSFLRPNDLVWNYVVGNYLKGnKPPPFD----
PhaC6        aVQLREmTLGEQAPnGPaLLKGQELATTFSFLRPNDLVWNYVVGNYLKGnKPPPFD----
PhaC3        aVQLREmTLGEQAPQGPaLLKGQELATTFSFLRPNDLVWNYVVGNYLKGnKPPPFD----
PhaC51       qVQLREaTIGldAPQGPGLLKGQELATTFSFLRPNDLVWNYVVGNYLKGDKPPPFD----
PhaC27       qVQLREaTIGldAPQGPGLLKGQELATTFSFLRPNDLVWNYVVGNYLKGDKPPPFD----
PhaC18       qVQLREaTIGldAPQGPGLLKGQELATTFSFLRPNDLVWNYVVGNYLKGDKPPPFD----
PhaC_P8N7    tlQLvEqisaDk-----GyLdGrymAaTFnlLRgrDLiWsYVtnNYLmGEdyaPFD----
PhaC_P2N8    qIaalEgkMa-----GaGyLdsaDmATaFnlLRaNDLVWsffVnNYLlGKdPaPFD----
```

```
                           391                         445
PhaC2₁₆      farWyehvvdLPGrMYleavkHlfklNrLar-GEfvalGERisLksIavPVYllAgadDd
PhaC2₅₀      fasWyehvvdLPGrLYleavkHlfklNrLar-GEfvalGERisLksItvPVYllAgadDd
PhaC_P2N10*  LvaWggDypdqaGraWlefvtacYreNrLlk-GtLvldGhtVDLrRlrlPVlnvfaRaDH
PhaC1₁₆      vLYWNSDSTrLPaaLHhdfc-dmvakNELgngevyqlpdaqVDLsRvDvesfIvAgRtDH
PhaC1₅₀      vLfWNSDaTrLPaaLHrdfc-ellktNELgrgevyqVedaKVDLrKvDvdsfVvAgRtDH
PhaC5        iLaWsvDgTNLPGkLHgqfL-difeaNaLpkkGsinIlGkplDLesIkvetlVtggttDH
PhaC20       vLYWNnDtTrLParLHgefL-dlfLlNpfRnPdaLellGkpiDvanvscdaYVvAgttDH
PhaC10*      ------------------------------------------------------------
PhaC2₁       iLYWNnDnTrLPaaLHgdlL-dffkhNpLshPGgLeVCGtpiDLqKvNvdsfsvAginDH
PhaC7*       ------------------------------------------------------------
PhaC_P11N2   LLYWNSDSTNLPGPfltWYLRnmYLeNnLRiPGKLamCGvKaDLGhvDmPsfVvAcREDH
PhaC25       LLYWNSDaTNLPGPfYtWYLRnTYheNkLakPnaLTVCGEKiDLGKlDiPaYIYgSREDH
PhaC_P1N3    LLYWNSDaTNLPGPMYCWYLRnTYheNkLavPGaLTVaGEqVDLGRIEAPVYlYgSREDH
PhaC14       LLYWNgDSTNLPGPMYCWYLRHTYLqNELRqPGKLTVCGEKiDLGaIDAPVYIYgSREDH
```

45

```
PhaC8        LLYWNgDSTNLPGPMYCWYLRHTYLtNELRqPGKLTVCGEKiDLGaIDAPVYIYgSREDH
PhaC23       LLYWNgDSTNLPGPMYCWYLRHTYLtNELRqPGKLTVCGEKiDLGaIDAPVYIYgSREDH
PhaC21       LLYWNgDSTNLPGPMYCWYLRHTYLtNELRqPGKLTVCGEKiDLGaIDAPVYIYgSREDH
PhaC4        LLYWNSDSTNLPGPflCWYLRnTYLeNkLKdPGavTVCGEKVDLGaIEAPVYlYsSREDH
PhaC_P4N10   LLYWNSDSTNLPGPflCWYLRnTYLeNkLKdPGavTVCGEKVDLGaIEAPVYlYsSREDH
PhaC19       LLYWNSDSTNLPGPMFCWYLRHTYLqNELRvPGKLvVCGEKVDLGKIkAPVfVYgSREDH
PhaC6        LLYWNSDSTNLPGPMFCWYLRHTYLqNELRvPGKLvVCGEKVDLGKIkAPVfVYgSREDH
PhaC3        LLYWNSDSTNLPGPMFCWYLRHTYLqNELRvPGKLvVCGEKVDLGKIkAPVfVYgSREDH
PhaC51       LLYWNgDSTNLPGPMFCWYLRHTYLsNELRqPGKLTVCGQKVDLGKIkAPVfVYASREDH
PhaC27       LLYWNgDSTNLPGPMFCWYLRHTYLnNELRqPGKLTVCGQKVDLGKIkAPVfVYASREDH
PhaC18       LLYWNgDSTNLPGPMFCWYLRHTYLnNELRqPGKLTVCGQKVDLGKIkAPVfVYASREDH
PhaC_P8N7    LLhWNSDtTNLPakwHraYLRdfYrdNkLvraGELvVdGtpiDihKvktPtYVqAgREDH
PhaC_P2N8    LLfWNaDaTrmPaaMHsfYLRnmYLhNrLRePGgiTlaGvpinLskITvPcYfvAtvEDH
```
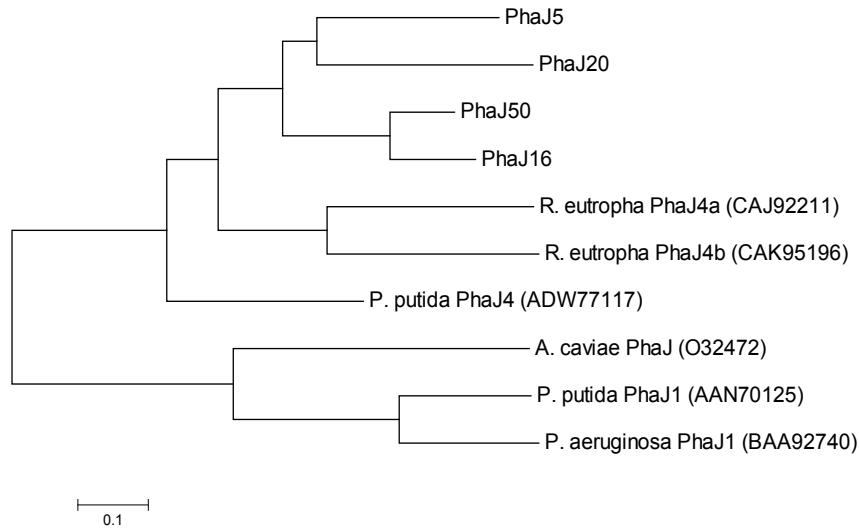
473 478

```
PhaC2(16)    IttpdqvfnaenLiGTPKdqiRkeLapgGHIglfmgaktlKqq-----------------
PhaC2(50)    IttpeqvfnaenLvGTPRdqiRkeLapgGHIglfmgaktlneq-----------------
PhaC_P2N10*  lVPaaAsqalgRai---gsrryqeLaieGghvGV--------------------------
PhaC1(16)    ItPWdAcYKSvnLL---gGKsqFVLvnSGHIqtlvcPPgKgKaSyqT--aDvLPeDphEW
PhaC1(50)    ItPWdgcYKSvnLL---gGesqFVLvnSGHIqtlvcPPgKgKaSyqT--gEtLPeDphEW
PhaC5        ltPWkgcYrtTQLL---gGpstFVLsnaGHvAslvNPPgnpKatyWl--gpKpddDpEaW
PhaC20       ItPWktcYatTQLL---tGereFVLsSSGHIqsIvNPPgnpKarflT--NpsfPaDpDDW
PhaC10*      ------------------------------------------------------------
PhaC2(1)     ItPWdAvYrSTlLL---gGerRFVLsnSGHvqsIlNPPsnpKanfve--stKLsgDpraW
PhaC7*       ------------------------------------------------------------
PhaC_P11N2   IVPWrAsYrgrRLL---gGKsRFVLGASGHIAGVINPPAKgKRnyWvnqgDagkasADEW
PhaC25       IVPiggAYaSTQLL---pGKkRFVmGASGHIAGVINPPAKKKRSHWiRddgKfPktqaEW
PhaC_P1N3    IVPiggAYaSTQLL---KGKkRFVmGASGHIAGVINPPAKKKRSHWiRadgKLPktqaEW
PhaC14       IVPWhAAYESTtiL---ntKrRFVLGASGHIAGVINPaAKnKRSHWv--NDKLPpiADEW
PhaC8        IVPWhAAYESTtiL---nsKrRFVLGASGHIAGVINPasKnKRSHWT--NDKLPasADDW
PhaC23       IVPWhAAYESTtiL---nsKrRFVLGASGHIAGVINPasKnKRSHWT--NDKLPasADDW
PhaC21       IVPWqAAYESTtiL---nsKrRFVLGASGHIAGVINPasKnKRSHWT--NDKLPpsADDW
PhaC4        IVPWtgAYaSTaLL---RGKkRFVLGASGHIAGVINPPAKKKRSHWT--NtttPkDpqrW
PhaC_P4N10   IVPWtgAYaSpaLL---RGKkRFVLGASGHIAGVINPPAKKKRSHWT--NtttPkDpqrW
PhaC19       IVPWqAAfESTKim---KGKvRFVLGASGHIAGVINPPaAKKRSHWv--dgKLgatAqQW
PhaC6        IVPWqAAfESTKim---KGKvRFVLGASGHIAGVINPPaAKKRSHWv--dgKLgatAqQW
PhaC3        IVPWqAAfESTKim---KGKvRFVLGASGHIAGVINPPaAKKRSHWv--dgKLgatAqQW
PhaC51       IVPWdAAfESTKiL---KGKvRFVLGASGHIAGVINPPaAKKRSHWa--ggKVdgDAqQW
PhaC27       IVPWdAAfESTKiL---KGKvRFVLGASGHIAGVINPPaAKKRSHWv--ggKVggDAqQW
PhaC18       IVPWdAAfESTKiL---KGKvRFVLGASGHIAGVINPPaAKKRSHWv--ggKVggDAqQW
PhaC_P8N7    IaPaasvwKiThyf---QGplRFVLagSGHIAGVvNPPeaQKyqyWT--NDgkaetlDQf
PhaC_P2N8    IaPWrsvYagaRLp---gGKtRFVLsgSGHIAGIINPPAahKyqHWT--cDsLPtspDEW
```

```
PhaC2(16)    -----------WariSAWivgqAa------------------------------------
PhaC2(50)    -----------WariSAWivqqg-------------------------------------
PhaC_P2N10*  -------------------fAGR-------------hglaiVpPAisRfltAtsiKKpr
PhaC1(16)    LehsKattGSWWemWikWaserhGpkkPA--PKApGs-aRfpSlgaAPGeYVrA-------
PhaC1(50)    LeqsKptiGSWWdlWvkWasqrsGakrPA--PKApGs-aKfppIgaAPGdYVrt-------
PhaC5        LqeAtkhtGtWWevWvdWagkrsGRqklA--PKAlGN-atYKArEkAPGtYVmerA-----
PhaC20       aAgAtphtGSWWehWTdWLatrsGerkaA--PKSlGs-ERHqplaaAPGtYVhg-------
PhaC10*      ------------------------------------------------------------
PhaC2(1)     yydAKQvdGSWWTqWlgWiQersGaLket--hmAlGN-QnYppmEaAPGtYVrvr------
PhaC7*       ------------------------------------------------------------
PhaC_P11N2   LeTAsEvsGSWWplWAeWLggfgGKkVtA--rrrlGs-adfsplEPAPGRYVKeKA-----
PhaC25       LAgAtEqPGSWWTDWAqWLKgHAGKqVPA--PKAYGdGKKYKAIEPAPGRYVKArA-----
PhaC_P1N3    LAgAtEhPGSWWTDWSnWLKtHAGKqIPA--PKSYGkGsaYKAIEsAPGRYVKArA-----
PhaC14       fvSAtErPGSWWpDWSAWLKaqgGaMIaA--PKgYGN-KtfKAIEPAPGRYVKqKA-----
PhaC8        MASAiEkPGSWWSDWSAWLKtqgGpMVaA--PKgYGN-KKlKAIEPAPGRYVKqKA-----
PhaC23       MAgAiEkPGSWWSDWSAWLKtqgGpMVaA--PKgYGN-KmlKAIEPAPGRYVKqKA-----
PhaC21       MASAtEkPGSWWSDWSAWLKtqgGpMVaA--PKgYGN-KKlKAIEPAPGRYVKqKA-----
PhaC4        LAgAtEsPGSWWTDWAAWLaplsGKeIaA--PKAlGs-RKYKSIEPAPGRYVKqKA-----
PhaC_P4N10   LAgAtEsPGSWWTDWAAWLaplsGKeIaA--PKAlGs-RKYKSIEPAPGRYVKqKA-----
PhaC19       LdSAKEvPGSWWTDWSAWLKpHAGKqVPA--PKtYGN-KsHKvIEPAPGRYVKAKA-----
PhaC6        LdSAKEvPGSWWTDWSAWLKpHAGKqVPA--PKtYGN-KsHKvIEPAPGRYVKAKA-----
PhaC3        LdSAKEvPGSWWTDWSAWLKpHAGKqVPA--PKtYGN-KsHKvIEPAPGRYVKAKA-----
PhaC51       LeSAKDvPGSWWSDWAAWLKpHAGKMVPA--PKAYGd-RsHKAIEaAPGRYVKAKA-----
PhaC27       LeSAKDvPGSWWSDWAAWLKpHAGKMVPA--PKAYGd-RsHKAIEaAPGRYVKAKA-----
PhaC18       LeSAKDvPGSWWSDWAAWLKpHAGKMVPA--PKAYGd-RsHKAIEaAPGRYVKAKA-----
PhaC_P8N7    vAgAtEhkGSWWpDWieWigaqdGKrVPAkgarvpGk-gKlKAIEnAPGsYVKAr------
PhaC_P2N8    qAqAaahaGSWWpDWkfWLgrrAGgkVmA---rtpGs-gglTAIEdAPGsYVrmtshQE--
```

Supplementary 3

A.

```
                              PhaJ5
                              PhaJ20
                              PhaJ50
                              PhaJ16
                              R. eutropha PhaJ4a (CAJ92211)
                              R. eutropha PhaJ4b (CAK95196)
                              P. putida PhaJ4 (ADW77117)
                              A. caviae PhaJ (O32472)
                              P. putida PhaJ1 (AAN70125)
                              P. aeruginosa PhaJ1 (BAA92740)
```

0.1

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-------GrlLvkarhTIEIEGESKPALIAEmLaMlItg-------
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