METAGENOMIC CHARACTERIZATION OF *CANDIDATATUS* DEFLUVIICOCCUS TETRAFORMIS TFO71, A TETRAD-FORMING ORGANISM, PREDOMINANT IN AN ANAEROBIC-AEROBIC MEMBRANE BIOREACTOR WITH DETERIORATED BIOLOGICAL PHOSPHORUS REMOVAL

BY

MASARU KONISHI NOBU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Environmental Engineering in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign 2013

Urbana, Illinois

Adviser:

Professor Wen-Tso Liu, Director of Research

ABSTRACT

In an acetate-fed anaerobic-aerobic membrane bioreactor with deteriorated enhanced biological phosphorus removal (EBPR), *Defluviicoccus*-related tetrad-forming organisms (DTFO) were observed to predominate in the microbial community. Using metagenomics, a partial genome of the predominant DTFO, "*Candidatus* Defluviicoccus tetraformis TFO71," was successfully constructed and characterized. Examining the genome confirmed the presence of genes related to the synthesis and degradation of glycogen and polyhydroxyalkanoate (PHA), which function as energy and carbon storage compounds. Both TFO71 and *Candidatus* Accumulibacter phosphatis (CAP) UW-1, a representative polyphosphate-accumulating organism (PAO), have PHA metabolism-related genes with high homology, but TFO71 has unique genes for PHA synthesis, gene regulation, and granule management. We further discovered genes encoding DTFO polyphosphate (polyP) synthesis, suggesting that TFO71 may synthesize polyP under untested conditions. However, TFO71 may not activate these genes under EBPR conditions because the retrieved genome does not contain inorganic phosphate transporters that are characteristic of PAOs (CAP UW-1, *Microlunatus phosphovorus* NM-1, and *Tetrasphaera* species). As a first step in characterizing EBPR-associated DTFO metabolism, this study identifies important differences between TFO and PAO that may contribute to EBPR community competition and deterioration.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Enhanced biological phosphorus removal (EBPR) systems use cyclic anaerobic-aerobic conditions to selectively enrich polyphosphate (polyP)-accumulating organisms (PAO) to remove phosphate (P_i) in discharged wastewater. PAO are characterized by their ability to aerobically synthesize polyP from soluble P_i as an energy source for anaerobic synthesis of polyhydroxyalkanoates (PHA) and/or glycogen from carbon substrates, which reciprocally contribute energy to aerobic polyP synthesis (Mino et al., 1998; He and McMahon, 2011a; He and McMahon, 2011b). Representative PAO include "*Candidatus* Accumulibacter phosphatis UW-1" (CAP UW-1), *Microlunatus phosphovorus* NM-1 and *Tetrasphaera* spp. (Nakamura et al., 1995a; Hanada et al., 2002; Kristiansen et al., 2012). Although the ability to cyclically store energy and survive anaerobic-aerobic environmental shifts allows PAO to dominate in EBPR systems, EBPR activities can deteriorate. This often coincides with a microbial community shift from PAO to "glycogen-accumulating organisms" (GAO) that can cyclically accumulate PHA and glycogen to survive and compete in the cyclic EBPR ecosystem (Fukase et al., 1985; Cech and Hartman, 1990, 1993; Liu et al., 1996).

Two types of GAO are often observed in full- and laboratory-scale EBPR systems: "*Candidatus* Competibacter phosphatis" of the class *Gammaproteobacteria* (Crocetti et al., 2002; Kong et al., 2002) and *Defluviicoccus*-related tetrad-forming organisms (DTFO) of the *Rhodospirillaceae* family in *Alphaproteobacteria* (Wong et al., 2004; Wong and Liu, 2006; McIlroy and Seviour, 2009; Mielczarek et al., 2013). Many studies have focused on ecophysiological characterization of DTFO as an important PAO competitor in EBPR systems (Dai et al., 2007; Wong and Liu, 2007; McIlroy and Seviour, 2009; McIlroy et al., 2010). These

studies characterized four *Defluviicoccus*-related GAO clusters (both tetrad-forming coccoid and filamentous morphology) (Wong et al., 2004; McIlroy and Seviour, 2009; McIlroy et al., 2010), and revealed valuable *in-situ* ecophysiological characterization of DTFO. For example, DTFO are thought to alternate glycogen and PHA accumulation during EBPR cycling and also perform fermentation; however, we require further detailed insight into their physiological and metabolic capabilities to obtain specific understanding of DTFO and PAO competition.

To improve our understanding of EBPR deterioration, it is important to understand GAO physiology and metabolic capabilities. While characterization of a DTFO isolated from an EBPR plant, *Defluviicoccus vanus* Ben114, certainly improved our comprehension of DTFO behavior (Maszenan et al., 2005; Wong and Liu, 2007), explicit characterization of DTFO specifically predominant in deteriorated in EBPR systems remains to be accomplished. Isolation of such DTFO has been unsuccessful because we cannot culture the majority of microorganisms with our current cultivation techniques (Amann et al., 1995; Puspite et al., 2012). Investigation of a PAO often associated with EBPR, CAP, also encountered difficulties due to cultivation issues. In order to circumvent this, previous studies implemented metagenomic approaches and achieved the great feat of characterizing physiological and metabolic mechanisms for how PAO accomplish EBPR (Garcia Martin et al., 2006; He and McMahon, 2011a). To complement these studies in defining EBPR microbial ecology, we set out to perform the first GAO metagenomic investigation and define DTFO metabolism and physiology.

Using modern metagenome construction and binning technology, we built a draft genome of "*Candidatus* Defluviicoccus tetraformis TFO71" from a metagenome of our previously reported cluster I DTFO-dominated microbial community of a laboratory-scale non-phosphorus-removing anaerobic-aerobic sequencing membrane bioreactor. (Wong and Liu, 2007). We

further metagenomically characterized TFO71's glycogen, PHA, and polyP metabolism and compared how it differs from representative PAO with full genome sequences (CAP UW-1 and *M. phosphovorus* NM-1) and partial genomes (*Tetrasphaera* spp.) (published by Kristiansen et al., but most genomic information still publically unavailable). With this characterization, we provide insight into differences between DTFO and PAO metabolism that may contribute to DTFO competitive advantage in EBPR systems.

CHAPTER 2

RESULTS AND DISCUSSION

2.1 - General Characteristics of the DTFO-dominated EBPR community metagenome

Results of 16S rRNA gene clone library and qualitative fluorescence *in-situ* hybridization (FISH) suggested that DTFO (>78% of total clones) were predominant in the reactor (Fig. 1, Fig. S1). The metagenome (IMG OID: 2199352010) constructed for this DTFO-predominated reactor community contained 19,897 scaffolds between 439 and 235,525 bp, totaling to 29.5 Mbp (Table 1). These scaffolds had a mean G+C content of 62% and encoded 43,464 proteins, of which 28,470 were annotated with predicted function.

We binned these scaffolds using PhyloPythia (McHardy et al., 2007) and read coverage- and G+C content (Droge and McHardy, 2012). While Phylopythia generated one predominant bin (11.2 Mbp) as expected, the average coverage of each scaffold varied between 2.4 and 2080.2 (Fig. 2). Given that sequencing coverage should loosely reflect relative abundance in the microbial community, this bin probably included sequences from both the target cluster I DTFO and other *Alphaproteobacteria*-related community members (Fig. 1). This bin contained two observable clusters separated by scaffolds' average read coverage: low (9.3±3.6) and high (360.7±300.6) coverage with G+C content of approximately ~66% (Fig. 2). The high coverage cluster contained a scaffold encoding a 16S rRNA gene (TFO_HY_00283270) (locus tags will be abbreviated by removing the "HY_00", e.g. – TFO_HY_00283270 will be abbreviated as TFO_283270) with 97.0% sequence homology to *Defluviicoccus vanus* Ben 114 and 97.7-98.6% to previously identified cluster I DTFO in activated sludge processes (AY351640, AY351628, and AY351639) (Wong et al., 2004) (Fig. 1). On the other hand, the low coverage cluster

contained a scaffold (average coverage of 43.7) encoding a 16S rRNA gene related to *Rhodobacteraceae* (99% similarity with *Rhodobacter* sp. Bo10-19). This suggests the high coverage cluster scaffolds were correctly binned to the predominant cluster I DTFO, while low coverage scaffolds belonged to other *Alphaproteobacteria*-related community members. We also manually binned high coverage (119.4~2380.5) short scaffolds (<1.5kb) corresponding to the cluster I DTFO bin, which Phylopythia may have missed due to the limitations of kmer frequency-based scaffold characterization (Fig. 2) (McHardy et al., 2007; Droge and McHardy, 2012). We confirmed that this bin possessed no more than one of each bacterial essential single copy genes proposed by Dupont *et al.* (Dupont et al., 2012). This binning yielded an approximately 92% complete 4.6 Mbp draft *Ca.* D. tetraformis TFO71 genome of composed of 56 RNA-coding genes and 4,165 predicted protein coding regions (77.8% with predicted function) and 56 RNA genes in 162 scaffolds with a mean G+C content of 64±4% (Table 1, supplemental results).

2.2 - Description of general metabolism

For central metabolism, TFO71 is capable of utilizing the Embden-Meyerhoff pathway (EMP), the tricarboxylic acid cycle (TCA), a partial pentose phosphate pathway, and complementary anaplerotic reactions (Tables S1 and S2). TFO71 also had the genetic capability to fix nitrogen (Table S3). As for substrate metabolism, the TFO71 genome contains no genes associated with chemolithotrophy and appears strictly organotrophic (Table S1, Table S2). In particular, the genome possesses genes for degradation of sugars (glucose, fructose, sucrose, mannose, maltose, lactose, cellulose, and glycogen), organic acids (acetate, lactate, pyruvate, and propionate), ethanol, and amino acids. To complement this organotrophy, TFO71 also had genes

encoding respiratory and fermentative pathways (Table S3). For respiration, TFO71 had genes for using oxygen as an electron acceptor. Although TFO71 had genes for nitrate reduction to ammonia and N_2, the nitrate reductase (NasA) is associated with assimilatory nitrate reduction rather than respiration (i.e., no denitrification) (Table S3). No genes for metal and sulfate respirations were found. The TFO71 genome suggested capability for lactate, propionate (methylmalonyl-CoA pathway), and succinate fermentation (Table S2).

2.3 - Glycogen synthesis and metabolism

GAO, including DTFO, are known to aerobically store glycogen as an energy source for the anaerobic stage of EBPR (Liu et al., 1996; Wong and Liu, 2007). TFO71's genome contained the complete set of genes for glycogen metabolism: glucose-ADP synthesis (*glgC*), glycogen synthesis (*glgA*), polymer organization (*glgB* and *glgE*), and degradation (*glgP* and *glgX*) (Fig. 3, Table S6). This confirms TFO71's ability to store and metabolize glycogen in DTFO-dominated deteriorated EBPR bioreactors (Wong et al., 2004; Wong and Liu, 2006, 2007). TFO71 possessed multiple gene cassettes associated with glycogen metabolism (Fig. S3, Table S6). All cassettes encode GlgA and associate it with various components of glycogen metabolism. Notably, while GlgA is necessary for aerobic glycogen synthesis, it associates with anaerobic glycogen degradation in most of the cassettes. Of the three GlgA, two are *Alphaproteobacteria*-related and one is *Verrucomicrobia*-related (TFO_278210) (Table S6). The reason for TFO71's redundancy in glycogen-related genes is unclear.

The draft genome indicates that TFO71 can ferment depolymerized glycogen or glucose to acetate, propionate, lactate, ethanol, and succinate. Balancing oxidation and reduction of metabolic intermediates is essential for cluster I DTFO survival during the anaerobic stage. For

this, we observed that acetate, fumarate, and pyruvate generated from glycolysis and TCA can serve as electron sinks for TFO71. The TFO71 genome further encodes hydrogenase subunits (TFO_275380 – 410) for H^+ respiration to complement fermentation as an alternative electron disposal pathway. However, under alternating anaerobic-aerobic conditions, we speculate that, rather than producing H_2, TFO71 uses the glucose-derived reducing power for converting fatty acids into PHA, which is explained in the following section.

2.4 - Polyhydroxyalkanoate (PHA) synthesis and metabolism

DTFO's anaerobic metabolism for synthesizing 3HA (PHA monomer) from stored glycogen and fed substrates is of particular interest because GAO and PAO have known differences in PHA synthesis and substrate integration into PHA that may influence competition: 3HA composition, substrate uptake rates, and rates of adaptation to substrate changes (Liu et al., 1997; Oehmen et al., 2005; Oehmen et al., 2006; Dai et al., 2008). Although our previous studies revealed that DTFO can survive EBPR conditions by cyclically synthesize/degrade PHA when fed acetate, propionate, lactate, or pyruvate (Wong and Liu, 2007), detailed investigation of the underlying mechanisms has not been possible because we lacked insight into DTFO PHA metabolism. We confirmed cluster I DTFO's capability to (de)polymerize PHA in two PHA depolymerase (PhaZ) (TFO_219210 and TFO_266630) and two class III PHA synthases (PhaCE) (TFO_278240 and TFO_278220), adapted to polymerizing C3-C5 3HA reflecting previously observed DTFO PHA composition (Nomura and Taguchi, 2007; Dai et al., 2008) (Fig. S4, Table S7). Further, TFO71's genome also encodes an acyl-CoA synthase, lactate dehydrogenase, and pyruvate dehydrogenase for assimilating/degrading these substrates (Fig. 3, Table S1, Table S2).

TFO71's genome encodes the metabolic capability to synthesize 3HA by condensing and reducing simple fatty acids (ACoA: acetyl-CoA, PCoA: propionyl-CoA, and MCoA: malonyl-CoA) generated from glycogen fermentation, substrate assimilation, and substrate degradation. ACoA acetyltransferase (*phaA*) and 3HA dehydrogenase (*phaB*) synthesizes 3HB (3-hydroxybutyrate); MCoA-ACP transacylase (*fabD*), acyl carrier protein (ACP), β-ketoacyl-ACP synthase II (*fabF*), and β-ketoacyl-ACP reductase (*fabG*) synthesizes 3HV (3-hydroxyvalerate), 3HMV (3-hydroxy-2-methylvalerate), and 3HMB (3-hydroxy-2-methylbutyrate); and enoyl-CoA hydratase (*phaJ*) synthesizes long-chain 3HA from enoyl-CoA produced from fatty acid beta-oxidation (Tsuge et al., 2003), a previously unobserved pathway for cluster I DTFO (Fig. 3, Fig. S4, Table S7). The Fab pathway particularly influences 3HA composition because these genes synthesize three of the four major 3HA. TFO71's Fab operon encodes two *fabF*s: while FabF1 (TFO_271560) has a conventional malonyl-ACP binding site, FabF2 (TFO_271590) is missing a binding site residue (Fig. S2) (Wang et al., 2006). Different binding site structure indicates that these FabF may also have different acyl-ACP affinities; whereby, inducing preferential synthesis of particular 3HA. For example, biased affinity for propionyl-ACP would increase 3HMB and 3HMV synthesis. This suggests that TFO71 may coordinate two FabF's to synthesize 3HV, 3HMB, and 3HMV.

While previous ecophysiological studies provided essential characterization of cluster I DTFO 3HA synthesis, PHA polymerization, and PHA degradation, understanding on the biological mechanisms that regulate these pathways in these organisms is still lacking. The TFO71 genome encodes two *Rhodospirillales*-specific PHA granule-associated proteins, phasins (PhaP), (TFO_276340 – 276330) (Table S7), which are thought to increase PHA biosynthesis, PHA granule number, and granule surface/volume ratio, and activate PHA depolymerization

(Handrick et al., 2004). To complement this PHA granule regulation, TFO71 also possesses a *Rhodospirillales*-specific PHA synthesis repressor (*phaR*) (TFO_271300) next to a transcriptional regulator (TFO_271310) that is not only highly conserved amongst PHA-synthesizing *Rhodospirillales,* but also often associated with *phaR* in *Rhodospirillales* genomes (see supplemental results). However, the role of this transcriptional regulator is unclear. Overall, we believe that TFO71 regulates PhaC-mediated PHA synthesis using PHA granule management (PhaP) and transcriptional regulation (PhaR).

2.5 - Polyphosphate synthesis and metabolism

The GAO-associated deterioration of phosphate removal in EBPR is attributed to their inability to accumulate polyP (Cech and Hartman, 1990, 1993). In agreement, DTFO and *D. vanus* Ben114 have been shown to not accumulate excess polyP under aerobic-anaerobic cycling (Wong et al., 2004) or aerobic growth (Maszenan et al., 2005). Unexpectedly, our metagenomic analysis revealed that the TFO71 genome contains an exopolyphosphatase (*ppx*) (TFO_273110) for polyP hydrolysis, and polyphosphate:AMP phosphotransferase (*pap*) and polyphosphate kinase 1 (*ppk1*) (TFO_273120 – 273140) for ATP/ADP-mediated polyP polymerization and AMP/ADP-mediated polyP dephosphrylation, respectively (Fig. 4, Fig. S5, and Table S8) (Kameda et al., 2001). We further observed that 93% of the *Rhodospirillales* genomes available on the IMG database (covering 41 species) possess the polyphosphate metabolism genes (ppk, pap, and ppx). Therefore, from a phylogenetic perspective, it is not surprising that TFO71's genome possesses these genes. The *ppx* and *pap*/*ppk1* may have interfering regulation based on the gene organization (Toledo-Arana and Solano, 2010), suggesting that TFO71 may separate usage of these genes based on environmental condition. To complement polyP metabolism, the

TFO71 genome contained an ATP-binding cassette phosphate transporter system (Pst) and Na^+/Pi symporter (TFO_05330), but no inorganic phosphate transporter (PiT family), which could have been missed during sequencing (Table S8). However, it is still unclear whether polyP is relevant to DTFO survival in EBPR.

2.6 - Genomic comparisons between Candidatus D. tetraformis TFO71 and PAOs

To improve our understanding of the GAO-PAO competition in EBPR systems, we compared TFO71 and representative PAO's strategy for surviving anaerobic-aerobic cycling and metabolic mechanisms for accomplishing this lifestyle. From the TFO71 genome, we confirm previous community-based observations that cluster I DTFO can perform fermentation and cyclically synthesize glycogen and PHA using each other as energy sources for the synthesis of the other. Furthermore, they cannot respire nitrate and may synthesize polyP under unknown conditions. In comparison with TFO71, aside from the PAO-characteristic ability to accumulate excess polyP (He and McMahon, 2011b), CAP UW-1 shares a very similar metabolic strategy, *M. phosphorvorus* NM-1 cannot perform fermentation but can respire nitrate (Nakamura et al., 1995b; Nakamura et al., 1995a; Akar et al., 2006; Kawakoshi et al., 2012), and *Tetrasphaera* spp. cannot synthesize PHA and requires nitrate for anaerobic metabolism (Hanada et al., 2002; Kristiansen et al., 2012). Here, we investigate TFO71 and PAO genomic similarities that allow them to survive anaerobic-aerobic cycling and also genomic differences that contribute to PAO-GAO competition.

It can be speculated that the EBPR environment may have imposed particular selective pressures on PAO and GAO that resulted in similar genetic adaptations. Whole-genome BLAST revealed that, by gene amino acid sequence similarity, TFO71's genome is no more closely

related to CAP UW-1 (43±11%) than *Azoarcus* (42±11%), a closely related genus of *Rhodocyclaceae*, or *M. phosphovorus* NM-1 (38±9%) than *Nocardioides* (39±9%), a closely related genus of *Propionibacterineae*. However, many TFO71 PHA metabolism and central metabolism genes had top BLAST hits to CAP UW-1: TCA (Mdh), methylmalonyl-CoA pathway (MmcDC, MmcB, and PccB), pyruvate metabolism (PckA), fatty acid oxidation (Ccr), PHA synthesis (PhaJ, PhaC, and PhaE), and PHA depolymerizaiton (PhaZ) (Table S1, Table S2, and Table S7).

Genomic comparisons also revealed that TFO71 and CAP UW-1 both utilize *glgABCEX* for glycogen synthesis and metabolism, *phaAB*, *fabDFG*, and *phaJ* for 3HA synthesis, and class I and III PHA synthases for PHA polymerization. TFO71 and CAP UW-1 also share many glycogen and PHA metabolism genes with high amino acid sequences similarities (up to 72%) (Table S6 and Table S7). In addition, many of their PHA metabolism gene cassettes share similar organization (Fig. S4). Although not having high homology in polyP metabolism genes, TFO71 and CAP UW-1 organize their *ppk1* and *ppx* in a similar orientation, suggesting interfering regulation (Fig. S5).

Despite such genomic similarities, differences in PHA metabolism-related ecophysiological traits have been reported between DTFO and PAO. CAP is thought to adapt faster than GAO to substrate changes (Oehmen et al., 2005). When fed with acetate, cluster I DTFO integrates more propionate into PHA than CAP, which suggests different DTFO-PAO metabolic behavior (see supplemental results). *M. phosphovorus* NM-1 can synthesize PHA under both aerobic and anaerobic conditions (Akar et al., 2006). While we speculate that GAO-PAO differences in PHA synthesis pathways may influence 3HA synthesis patterns and substrate adaptation, the underlying metabolic mechanisms that contribute to these phenomena had not

been elucidated. As preliminary steps, we identify that TFO71 uniquely possesses a FabF with non-conventional binding site (Fig. S1) and a *Rhodospirillales*-related PhaP. The FabF2 may influence how TFO71 incorporates fed substrates and glycogen fermentation products into PHA (Fig. S6). For example, a higher propionate affinity would uniquely increase TFO71's 3HV and 3HMV synthesis, reflecting acetate-fed DTFO and CAP PHA composition (see supplemental results), and, thereby increase reductive glycogen degradation (Fig. S6). For PhaP, CAP UW-1 possesses a *Cupriavidus necator* H16-related PhaP non-homologous to that of TFO71 (Table S7, Table S9). Distinct from *C. necator* H16-type PhaP, *Rhodospirillales*-type PhaP stimulates both PHA polymerization and degradation (York et al., 2001; Handrick et al., 2004; Hanisch et al., 2006); therefore, TFO71 may also have a unique PHA metabolism scheme for alternating between anaerobic and aerobic survival.

While TFO71 and CAP UW-1 share similar overarching metabolic strategies for surviving anaerobic-aerobic cycling, *M. phosphovorus* NM-1 lacks many 3HA metabolism and PHA synthesis genes that we identify in TFO71's genome (*phaAB* and *phaCE* respectively) (Kawakoshi et al., 2012). Instead, NM-1 has acyl-CoA beta-oxidation genes (*yfxYX*) that are thought to oxidatively synthesize 3HA from acyl-CoA molecules (Kawakoshi et al., 2012). This indicates that NM-1 can synthesize PHA both oxidatively and reductively (*yfxYX* and *fabDFG* correspondingly), but TFO71 and CAP UW-1 can only synthesize PHA reductively. This reflects previous observations that *M. phosphovorus* NM-1 can both aerobically and anaerobically synthesize PHA unlike TFO71 or CAP (Akar et al., 2006). Furthermore, the *M. phosphovorus* NM-1 genome contains neither PhaP nor PhaZ homologs. Therefore, TFO71 and *M. phosphovorus* NM-1 clearly have different PHA metabolism/regulation and PHA-dependent survival strategies.

Currently, we lack insight into DTFO polyP metabolism because we have not been able to detect polyP accumulation in DTFO. Based on DTFO and PAO genomic comparisons, TFO71 encodes no PiT genes, and PAO genomes (CAP UW-1, *M. phosphovorus* NM-1, and *Tetrasphaera* spp.) have multiple copies (Kawakoshi et al., 2012; Kristiansen et al., 2012) (Table S9). Although the remaining non-sequenced TFO71 genome may encode PiT gene(s), 90% of *Rhodospirillales* genomes available on IMG that possess polyP metabolism genes encode no more than one copy of PiT. Thus, it is unlikely that TFO71's genome would encode more than one PiT. A previous study showed that CAP hydrolyzed polyP, accumulated P_i, used the high intracellular P_i to drive paired cation/P_i export through the PiT system, and used the generated proton (or cation) motive force to assimilate acetate for PHA synthesis (Saunders et al., 2007), suggesting that PiT may serve an essential role in PAO harnessing energy from aerobically synthesized polyP for anaerobic substrate uptake. DTFO may not require this function because polyP is not thought to be an essential compound for GAO transition from aerobic to anaerobic metabolism.

2.7 - Conclusion

In summary, our metagenomic analysis confirms that TFO71 can alternate PHA and glycogen storage during the EBPR anaerobic-aerobic cycling; in the anaerobic stage, TFO71 can convert glycogen into PHA through fermentation; and in the aerobic stage, TFO71 can degrade PHA through aerobic respiration and synthesize glycogen. The genetic and metabolic differences observed between DTFO and PAO further raise key questions: "Do DTFO accumulate polyP at low levels in EBPR systems to help their survival?", "Are PiT systems specifically necessary for polyphosphate metabolism?", "How does TFO71's FabF2 influence 3HA synthesis patterns and

fatty acid integration into PHA?", and "How does the novel coordination of *Rhodospirillales* PhaP influence regulation of PHA synthesis and degradation?" Given that DTFO and CAP UW-1 clearly have metabolic differences from previous molecular ecological studies and this metagenomic study, the next step is to reveal how DTFO- and PAO-specific genes contribute to such differences. This study identified such genes and provides targets for future gene expression analysis for DTFO-PAO competition in the EBPR systems. If these genes indeed contribute to competitive differences between DTFO and PAO, characterization of the function and regulation of these genes would be of utmost interest.

CHAPTER 3

DESCRIPTION OF *CANDIDATUS* DEFLUVIICOCCUS TETRAFORMIS

Defluviicoccus tetraformis (Gr.n. *tetra*, four; L. suff. *formis* [from L. n. *forma*, figure, shape, appearance], -like, in the shape of; N.L. fem. adj. *tetraformis*, tetrad-formed, referring to the morphology of bacterial cells).

Facultatively anaerobe, chemoheterotrophic. Cocci cells (1-2 μm) form clusters predominantly tetrads. Represent a novel species in the genus *Defluviicoccus* based on 16S rRNA gene sequence analysis. Enriched from an acetate-fed anaerobic-aerobic membrane EBPR bioreactor with deteriorated phosphorus removal.

CHAPTER 4

EXPERIMENTAL PROCEDURES

4.1 – Sample Collection

A sequencing membrane bioreactor (MBR) was operated under cyclic anaerobic and aerobic conditions with acetate as the sole carbon source as described previously (Wong and Liu, 2007). Sludge samples taken from MBR were fixed with ethanol and stored at -20˚C until use for fluorescence in situ hybridization analysis, clone library analysis, and metagenomic sequencing.

FISH

FISH was performed using a cluster I DTFO-targeting probe, TFO_DF776 (5'-AGGACTTTCACGCCTCAC-3'), as previously described (Amann et al., 1995; Wong and Liu, 2007). The probe was synthesized and labeled at 5' end with cyanine 3 (Cy3) (Integrated DNA Technologies, Inc., Iowa, USA). Olympus BX51 and Carl Zeiss LSM 5 Pascal were used for epifluorescence and confocal laser scanning microscopy observations correspondingly as previously described (Wong and Liu, 2007). Images were analyzed as in a previous study (Liu et al., 2001).

4.2 – DNA extraction and metagenomic sequencing

Total high molecular weight DNA was extracted from the sludge sample taken at 250 days using Schmidt's protocol and was stored at -80°C until use (Schmidt et al., 1991). The DNA obtained was directly sequenced using the Genome Sequencer GS FLX Titanium platform (Roche, Switzerland) and Illumina GAIIx (Illumina, San Diego, USA) at Roy J. Carver Biotechnology Center at the University of Illinois. Sequence quality trimming was done by

Quake using a minimum quality score of 20 and minimum length of 60 (http://www.cbcb.umd.edu/software/quake/) (Kelley et al., 2010).

4.3 – 16S rRNA gene clone library analysis

The genomic DNA used for metagenomic sequencing was subjected to the clone library analysis. 16S rRNA genes were amplified by PCR using the following universal primers: 8F (5'-AGAGTTTGATCMTGGCTCAG-3') (Tamaki et al., 2005) and 1492R (5'-TACGGYTACCTTGTTACGACTT-3'). The PCR mixture (50 μL) contained 1×PCR buffer, 3.5 mM MgCl2, 10 mM deoxynucleoside triphosphates (dNTPs), 1.25 U AmpliTaq Gold (each from Applied Biosystems, Foster City, CA, USA), and 0.4 μM of each forward and reverse primer. Approximately 100 ng of genomic DNA was used as a template under the following cycling conditions: initial AmpliTaq Gold activation at 95˚C for 9 min, followed by 20 cycles of denaturation (to limit amplification bias) at 95˚C for 30 s, annealing at 55˚C for 30 s, extension at 72˚C for 2 min, and a final extension step at 72˚C for 5 min. The purified 16S rRNA genes were cloned with a pT7Blue T-vector kit (Novagen, Madison, WI, USA). The clonal DNAs were amplified from randomly selected recombinants by direct PCR with M13 primers, and then used as templates for sequencing. A universal primer, 907r (5'-CCGYCAATTCMTTTRAGTTT-3'), was used for sequencing of the cloned 16S rRNA genes. The sequencing was conducted at Roy J. Carver Biotechnology Center at the University of Illinois. The obtained sequences of all the 16S rRNA gene clones (~951 bp) were compared with those in the GenBank database by using the BLAST program (NCBI-BLAST, www.ncbi.nlm.nih.gov/BLAST).

4.4 – Bioinformatics

Sequence data sets obtained by 454 FLX Titanium and Illumina GAIIx were assembled using Roche/454 De Novo Assembler (Newbler) (Margulies et al., 2005) and CLC Genomic Workbench (CLC bio, Denmark), respectively. Further hybrid assembly using both 454- and Illumina-contigs was performed with CLC Genomic Workbench. CLC and Newbler assembly were conducted with default settings except for the parameters of sequence similarity and overlapped sequence length (Newbler) or length fraction (CLC) as follows: 98% of sequence similarity and 50 bp of overlapped sequence length were used for Newbler; as for CLC assembly, sequence similarity and length fraction were set to 0.98 and 0.4.

All genes contained in the metagenomic scaffolds were identified and annotated using the Joint Genome Institute's Integrated Microbial Genomes and Metagenomes (IMG/M) pipeline (Markowitz et al., 2012). Through Phylopythia, we characterized scaffolds kmer frequency, used the largest scaffold as a reference, binned other scaffolds to this reference based on the built-in support vector machine classifier, and iterated this process with non-binned scaffolds ≥10kbp (McHardy et al., 2007). PhyloPythia can accurately categorize sequences ≥1-3kbp, but this can be highly dependent on sample community composition (McHardy et al., 2007). For the purposes of this study, we focused on the primary bin that presumably belonged to the cluster I DTFO based on its abundance in the reactor community. Inherently, kmer frequency-based binning has two limitations: incorrect binning of (i) closely related organisms with similar kmer frequency and (ii) short sequences that may have skewed kmer frequency caused by transposons or localized low-GC regions. To ensure accurate binning, the average coverage of each scaffold from the TFO71 Phylopythia bin was compared. Based on the skewed community structure (predominated by TFO71), we expect TFO71 scaffolds to have high coverage while scaffolds

belonging to other organisms have markedly lower coverage. Therefore, based on the distribution of the metagenomic scaffolds' coverage, we determined the range of TFO71 scaffold coverage. Scaffolds with coverage below the threshold were removed from the Phylopythia bin. In addition, to accommodate for Phylopythia's binning limitations for short sequences, shorter scaffolds that fell within the TFO71 coverage range were also added to the Phylopythia-based bin. We validated the assembly quality of these scaffolds using the Assembly Likelihood Estimator, and rejected any scaffolds with regions of markedly poor read quality, low read coverage, high paired-end insert length, or poor read alignment (Clark et al., 2013). The final bin was further checked for sequence contamination by verifying that the draft genome only contained one copy of each bacterial essential single copy genes proposed by Dupont *et al*. (Dupont et al., 2012).

REFERENCES

Akar, A., Akkaya, E.U., Yesiladali, S.K., Celikyilmaz, G., Cokgor, E.U., Tamerler, C. et al. (2006) Accumulation of polyhydroxyalkanoates by *Microlunatus phosphovorus* under various growth conditions. *J Ind Microbiol Biotechnol* **33**: 215-220.

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

Cech, J.S., and Hartman, P. (1990) Glucose-induced break down of enhanced biological phosphate removal. *Environ Technol* **11**: 651-656.

Cech, J.S., and Hartman, P. (1993) Competition between polyphosphate and polysaccharide accumulating bacteria in enhanced biological phosphate removal systems. *Water Res* **27**: 1219-1225.

Clark, S.C., Egan, R., Frazier, P.I., and Wang, Z. (2013) ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics* **29**: 435-443.

Crocetti, G.R., Banfield, J.F., Keller, J., Bond, P.L., and Blackall, L.L. (2002) Glycogen-accumulating organisms in laboratory-scale and full-scale wastewater treatment processes. *Microbiology* **148**: 3353-3364.

Dai, Y., Lambert, L., Yuan, Z., and Keller, J. (2008) Characterisation of polyhydroxyalkanoate copolymers with controllable four-monomer composition. *J Biotechnol* **134**: 137-145.

Dai, Y., Yuan, Z.G., Wang, X.L., Oehmen, A., and Keller, J. (2007) Anaerobic metabolism of *Defluviicoccus vanus* related glycogen accumulating organisms (GAOs) with acetate and propionate as carbon sources. *Water Res* **41**: 1885-1896.

Droge, J., and McHardy, A.C. (2012) Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief Bioinform* **13**: 646-655.

Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R. et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* **6**: 1186-1199.

Fukase, T., Shibata, M., and Miyaji, Y. (1985) Factors Affecting Biological Removal of Phosphorus. *Water Sci Technol* **17**: 187-198.

Garcia Martin, H., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C. et al. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24**: 1263-1269.

Hanada, S., Liu, W.T., Shintani, T., Kamagata, Y., and Nakamura, K. (2002) *Tetrasphaera elongata* sp nov., a polyphosphate-accumulating bacterium isolated from activated sludge. *Int J Syst Evol Microbiol* **52**: 883-887.

Handrick, R., Reinhardt, S., Schultheiss, D., Reichart, T., Schuler, D., Jendrossek, V., and Jendrossek, D. (2004) Unraveling the function of the *Rhodospirillum rubrum* activator of polyhydroxybutyrate (PHB) degradation: the activator is a PHB-granule-bound protein (phasin). *J Bacteriol* **186**: 2466-2475.

Hanisch, J., Walterrnann, M., Robenek, H., and Steinbuchel, A. (2006) The *Ralstonia eutropha* H16 phasin PhaP1 is targeted to intracellular triacylglycerol inclusions in *Rhodococcus opacus* PD630 and *Mycobacterium smegmatis* mc(2)155, and provides an anchor to target other proteins. *Microbiology* **152**: 3271-3280.

He, S., and McMahon, K.D. (2011a) '*Candidatus* Accumulibacter' gene expression in response to dynamic EBPR conditions. *ISME J* **5**: 329-340.

He, S.M., and McMahon, K.D. (2011b) Microbiology of '*Candidatus* Accumulibacter' in activated sludge. *Microbial Biotechnology* **4**: 603-619.

Kameda, A., Shiba, T., Kawazoe, Y., Satoh, Y., Ihara, Y., Munekata, M. et al. (2001) A novel ATP regeneration system using polyphosphate-AMP phosphotransferase and polyphosphate kinase. *Journal of bioscience and bioengineering* **91**: 557-563.

Kawakoshi, A., Nakazawa, H., Fukada, J., Sasagawa, M., Katano, Y., Nakamura, S. et al. (2012) Deciphering the genome of polyphosphate accumulating *Actinobacterium Microlunatus phosphovorus*. *DNA Res* **19**: 383-394.

Kelley, D.R., Schatz, M.C., and Salzberg, S.L. (2010) Quake: quality-aware detection and correction of sequencing errors. *Genome Biol* **11**.

Kong, Y.H., Ong, S.L., Ng, W.J., and Liu, W.T. (2002) Diversity and distribution of a deeply branched novel proteobacterial group found in anaerobic-aerobic activated sludge processes. *Environ Microbiol* **4**: 753-757.

Kristiansen, R., Nguyen, H.T., Saunders, A.M., Nielsen, J.L., Wimmer, R., Le, V.Q. et al. (2012) A metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological phosphorus removal. *ISME J*.

Liu, W.T., Mino, T., Nakamura, K., and Matsuo, T. (1996) Glycogen accumulating population and its anaerobic substrate uptake in anaerobic-aerobic activated sludge without biological phosphorus removal. *Water Res* **30**: 75-82.

Liu, W.T., Nakamura, K., Matsuo, T., and Mino, T. (1997) Internal energy-based competition between polyphosphate- and glycogen-accumulating bacteria in biological phosphorus removal reactors - Effect of P/C feeding ratio. *Water Res* **31**: 1430-1438.

Liu, W.T., Nielsen, A.T., Wu, J.H., Tsai, C.S., Matsuo, Y., and Molin, S. (2001) In situ identification of polyphosphate- and polyhydroxyalkanoate-accumulating traits for microbial populations in a biological phosphorus removal process. *Environ Microbiol* **3**: 110-122.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

Markowitz, V.M., Chen, I.M.A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y. et al. (2012) IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123-D129.

Maszenan, A.M., Seviour, R.J., Patel, B.K.C., Janssen, P.H., and Wanner, J. (2005) *Defluvicoccus vanus* gen. nov., sp nov., a novel Gram-negative coccus/coccobacillus in the '*Alphaproteobacteria*' from activated sludge. *Int J Syst Evol Microbiol* **55**: 2105-2111.

McHardy, A.C., Martin, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nat Methods* **4**: 63-72.

McIlroy, S., and Seviour, R.J. (2009) Elucidating further phylogenetic diversity among the *Defluviicoccus*-related glycogen-accumulating organisms in activated sludge. *Environmental Microbiology Reports* **1**: 563-568.

McIlroy, S.J., Nittami, T., Seviour, E.M., and Seviour, R.J. (2010) Filamentous members of cluster III *Defluviicoccus* have the in situ phenotype expected of a glycogen-accumulating organism in activated sludge. *Fems Microbiology Ecology* **74**: 248-256.

Mielczarek, A.T., Nguyen, H.T.T., Nielsen, J.L., and Nielsen, P.H. (2013) Population dynamics of bacteria involved in enhanced biological phosphorus removal in Danish wastewater treatment plants. *Water Res* **47**: 1529-1544.

Mino, T., Van Loosdrecht, M.C.M., and Heijnen, J.J. (1998) Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Res* **32**: 3193-3207.

Nakamura, K., Ishikawa, S., and Kawaharasaki, M. (1995a) Phosphate-uptake and release activity in immobilized polyphosphate-accumulating bacterium *Microlunatus phosphovorus* strain NM-1. *Journal of Fermentation and Bioengineering* **80**: 377-382.

Nakamura, K., Hiraishi, A., Yoshimi, Y., Kawaharasaki, M., Masuda, K., and Kamagata, Y. (1995b) *Microlunatus Phosphovorus* gen. nov, sp. nov, a new gram-positive polyphosphate-accumulating bacterium isolated from activated-sludge. *Int J Syst Bacteriol* **45**: 17-22.

Nomura, C.T., and Taguchi, S. (2007) PHA synthase engineering toward superbiocatalysts for custom-made biopolymers. *Appl Microbiol Biotechnol* **73**: 969-979.

Oehmen, A., Yuan, Z.G., Blackall, L.L., and Keller, J. (2005) Comparison of acetate and propionate uptake by polyphosphate accumulating organisms and glycogen accumulating organisms. *Biotechnol Bioeng* **91**: 162-168.

Oehmen, A., Saunders, A.M., Vives, M.T., Yuan, Z.G., and Keller, H. (2006) Competition between polyphosphate and glycogen accumulating organisms in enhanced biological phosphorus removal systems with acetate and propionate as carbon sources. *J Biotechnol* **123**: 22-32.

Puspite, I.D., Kamagata, Y., Tanaka, M., Asano, K., and Nakatsu, C.H. (2012) Are uncultivated Bacteria really uncultavable? *Microbes Environ* **27**: 356-366.

Saunders, A.M., Mabbett, A.N., McEwan, A.G., and Blackall, L.L. (2007) Proton motive force generation from stored polymers for the uptake of acetate under anaerobic conditions. *FEMS Microbiol Lett* **274**: 245-251.

Schmidt, T.M., Delong, E.F., and Pace, N.R. (1991) Analysis of a Marine Picoplankton Community by 16s Ribosomal-Rna Gene Cloning and Sequencing. *J Bacteriol* **173**: 4371-4378.

Tamaki, H., Sekiguchi, Y., Hanada, S., Nakamura, K., Nomura, N., Matsumura, M., and Kamagata, Y. (2005) Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. *Applied and Environmental Microbiology* **71**: 2162-2169.

Toledo-Arana, A., and Solano, C. (2010) Deciphering the physiological blueprint of a bacterial cell revelations of unanticipated complexity in transcriptome and proteome. *Bioessays* **32**: 461-467.

Tsuge, T., Taguchi, K., Seiichi, T., and Doi, Y. (2003) Molecular characterization and properties of (R)-specific enoyl-CoA hydratases from *Pseudomonas aeruginosa*: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid beta-oxidation. *Int J Biol Macromol* **31**: 195-205.

Wang, J., Soisson, S.M., Young, K., Shoop, W., Kodali, S., Galgoci, A. et al. (2006) Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* **441**: 358-361.

Wong, M.T., and Liu, W.T. (2006) Microbial succession of glycogen accumulating organisms in an anaerobic-aerobic membrane bioreactor with no phosphorus removal. *Water Sci Technol* **54**: 29-37.

Wong, M.T., and Liu, W.T. (2007) Ecophysiology of *Defluviicoccus*-related tetrad-forming organisms in an anaerobic-aerobic activated sludge process. *Environ Microbiol* **9**: 1485-1496.

Wong, M.T., Tan, F.M., Ng, W.J., and Liu, W.T. (2004) Identification and occurrence of tetrad-forming *Alphaproteobacteria* in anaerobic-aerobic activated sludge processes. *Microbiology* **150**: 3741-3748.

York, G.M., Stubbe, J., and Sinskey, A.J. (2001) New insight into the role of the PhaP phasin of *Ralstonia eutropha* in promoting synthesis of polyhydroxybutyrate. *J Bacteriol* **183**: 2394-2397.

TABLES

Table 1. Statistics of assembled metagenome and binned TFO71 genome.

	Metagenome	Binned (TFO71)
IMG Taxon object ID	2199352010	
DNA (bp)	29,528,629	4,641,237
GC (%)	62	64
Scaffolds	19,897	162
longest scaffold (bp)	235,525	235,525
Genes	43,825	4,221
- Protein-coding	43,464	4,165
w/ predicted function	28,470	3,240
w/o predicted function	14,994	925
- rRNA	54	3
- tRNA	307	53
% of total metagenome		
- by raw reads		80.4
- by assembled reads		0.8
- by sequence length		15.7

FIGURES

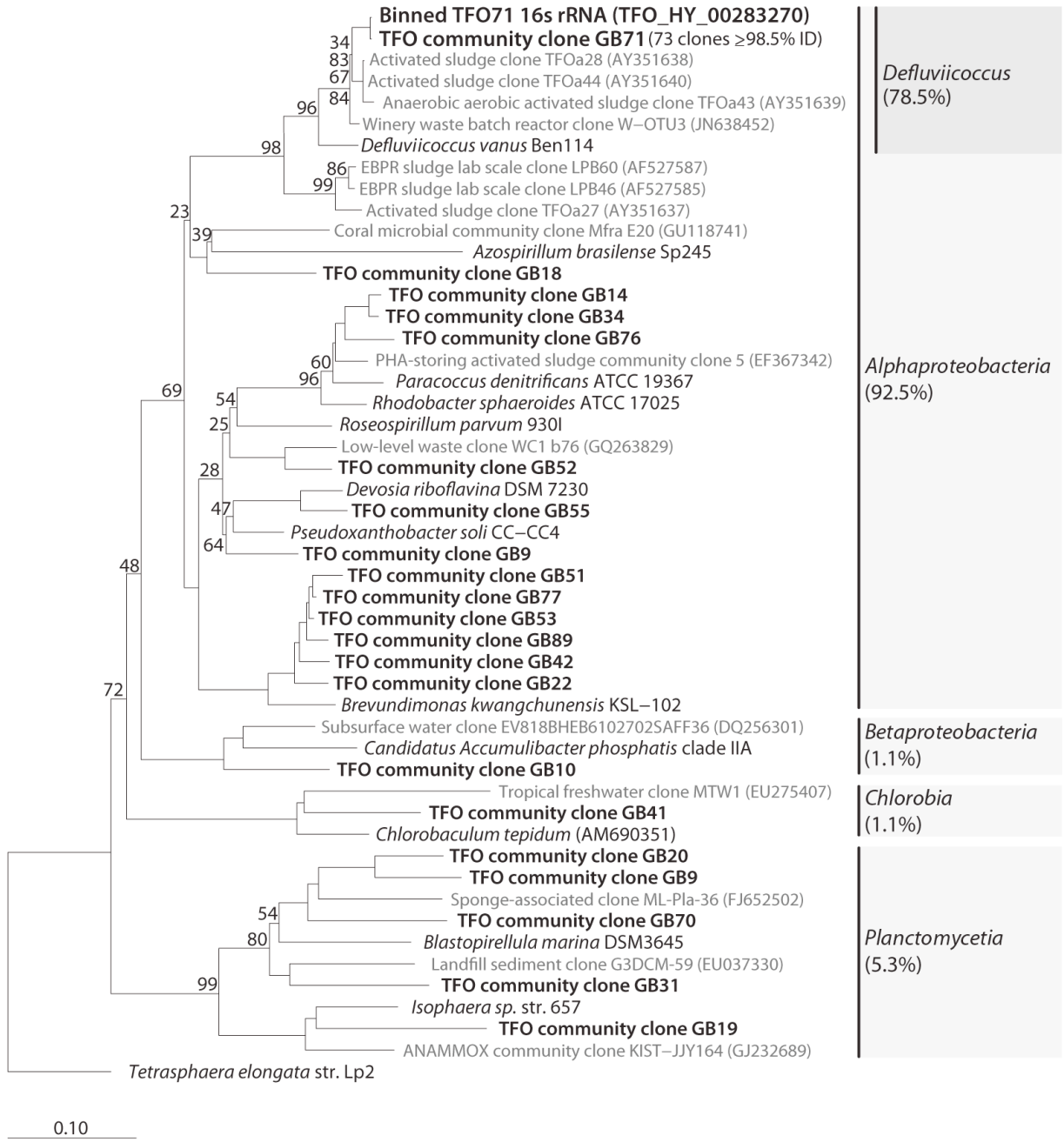


Figure 1. Phylogenetic tree of reactor DTFO community 16s rRNA clones. Bolded entries are 16s rRNA clones derived from the reactor community. Out of 93 clones, 73 (78.5%) shared greater than or equal to 98.5% sequence similarity and are grouped as clone GB71. The clone group GB71 falls into the *Defluviicoccus* genus of *Alphaproteobacteria*. At class level classification, *Alphaproteobacteria*, *Betaproteobacteria*, *Chlorobia*, and *Planctomycetia* composed 92.5%, 1.1%, 1.1%, and 5.3% of the community clones respectively. The 16s rRNA sequence found in binned TFO71 draft genome is closely related to the GB71 clone group.

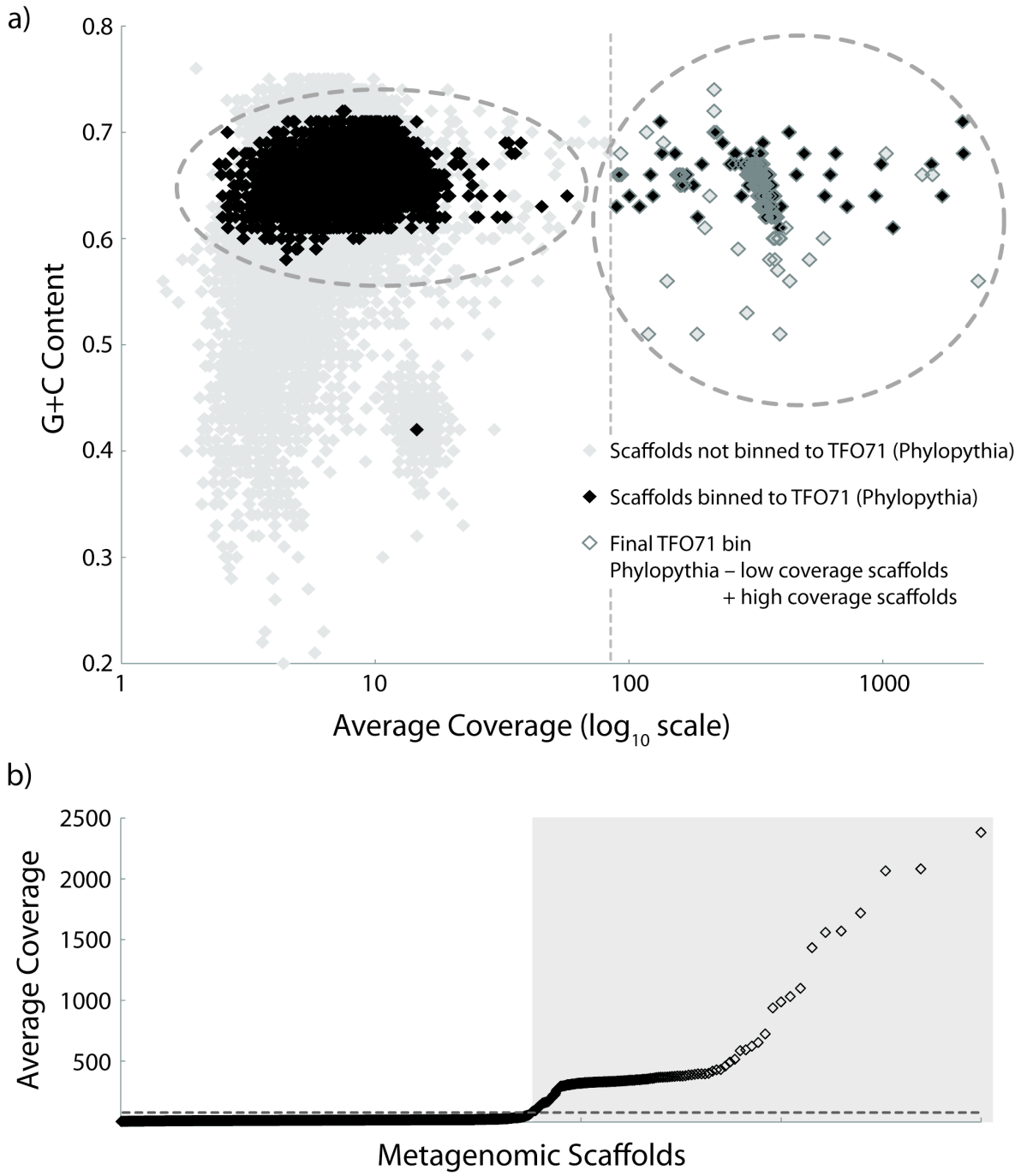


Figure 2. Metagenomic scaffolds plotted by average coverage (log-scale) and G+C% content. Scaffolds binned by Phylopythia (black diamonds) generally had G+C content ranging from 0.6 – 0.7 while scaffolds with higher or lower G+C content were mostly not binned (light gray diamonds). This Phylopythia bin was improved by filtering out low-coverage scaffolds (left dotted-line circle) and manually binning high-coverage short sequences (right dotted-line circle) missed by Phylopythia. The final TFO71 bin (dark gray bordered diamonds) had high coverage ranging from 89–2380 and a G+C content ranging from 0.50–0.74.

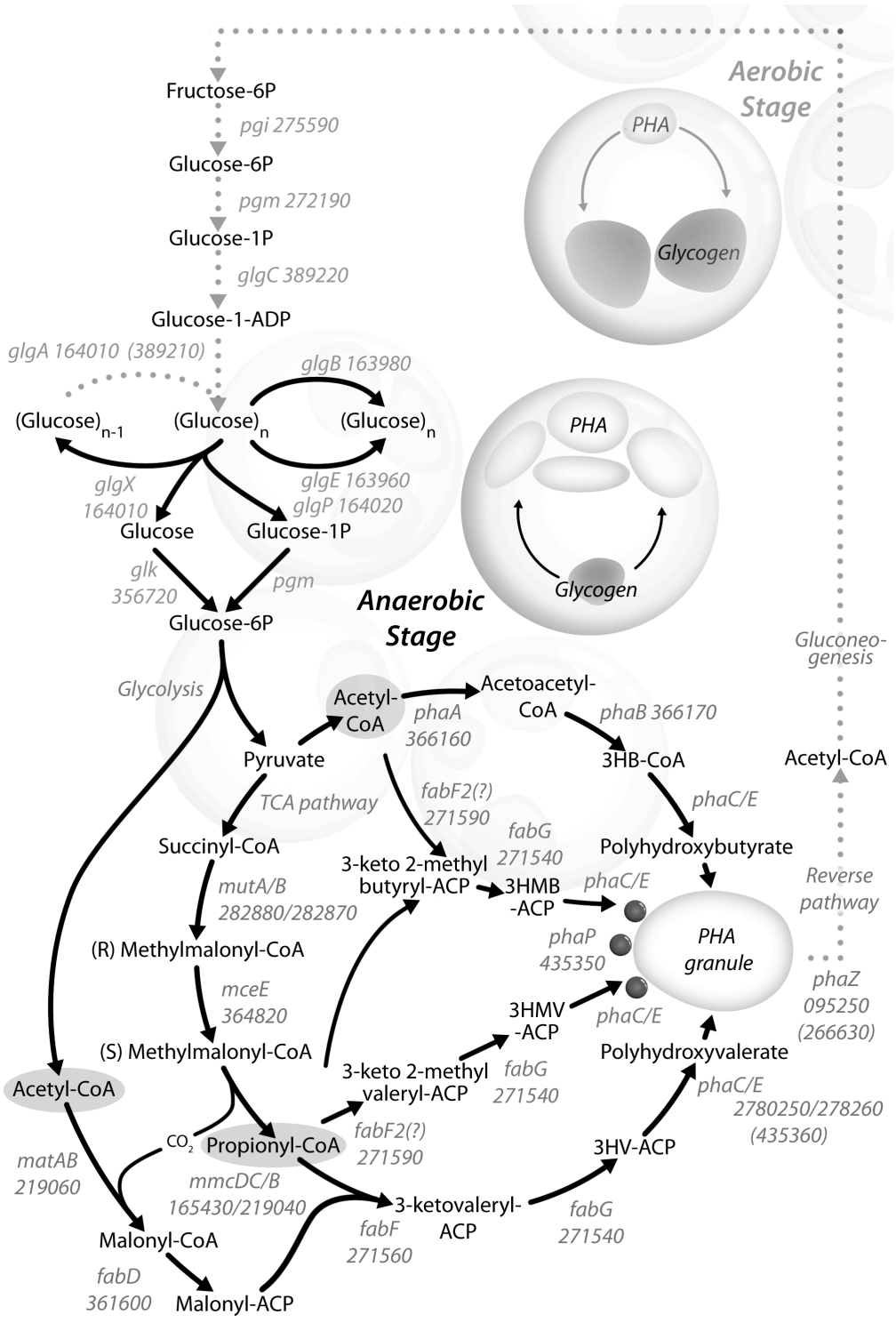


Figure 3. Metabolic pathways for glycogen and PHA metabolism and synthesis. Each reaction is labeled with corresponding TFO71 loci encoding the specific genes. Aerobic and anaerobic pathways are represented by gray and black arrows correspondingly. TFO71 locus tags are abbreviated (TFO_HY_00123456 = "123456"). Refer to supplemental tables for detailed information for each locus.

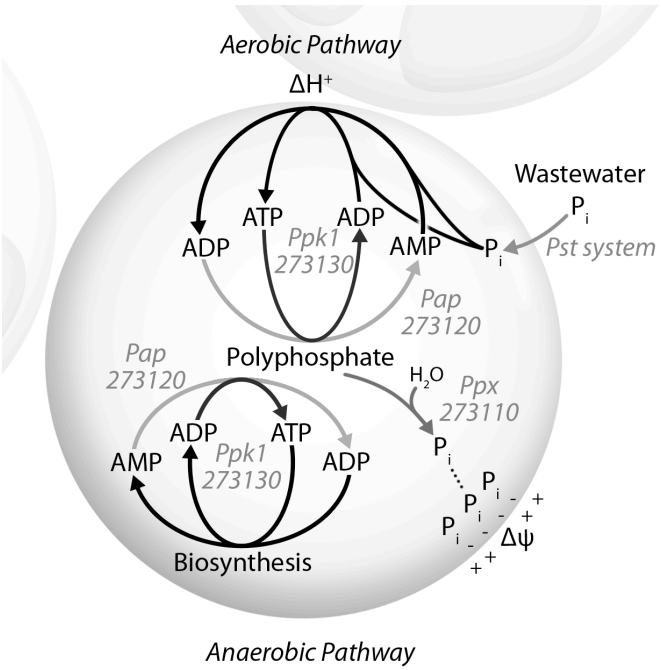


Figure 4. Metabolic pathways for polyphosphate metabolism and synthesis. Each reaction is labeled with corresponding TFO71 loci encoding the specific genes. The aerobic and anaerobic pathways are separated into top and bottom respectively. TFO71 locus tags are abbreviated (TFO_HY_00123456 = "123456"). Refer to supplemental tables for detailed information for each locus.

APPENDIX A

SUPPLEMENTAL RESULTS

Genome coverage

The draft genome contains 102 of 111 bacterial essential single copy genes proposed by Dupont et al. (Dupont et al., 2012). Genomes of members of the family *Rhodospirillales* only contain 110 of these genes. The TFO71 draft genome possesses 102 of 110 essential single copy genes for *Rhodospirillales*; therefore, we estimate that this draft genome has approximately 92.7% coverage.

Comparison of 3HA synthesis patterns in DTFO and CAP

Propionate-fed cluster I DTFO and CAP produce similar ratios of 3HMV and 3HV (~1:1) on a molar basis. On the other hand, when fed with acetate, they produce dissimilar ratios of 3HB and 3HV (3.1:1 and 5.8:1 respectively) and DTFO uniquely produces 3HMV (Oehmen et al., 2005; Dai et al., 2008). Therefore, DTFO apparently integrates more propionate into 3HA synthesis (3HV and 3HMV both require propionate where as 3HB does not).

Given the 3HA/PHA synthesis pathways (Lemos et al., 2003), 3HA synthesis patterns most likely influence how fed substrates (acetate and propionate) and glycogen fermentation products (ACoA and PCoA) are incorporated into PHA (Fig. S6). Furthermore, GAO and PAO must balance the oxidative and reductive production of ACoA and PCoA to generate the appropriate amount of reducing equivalents for PHA synthesis (Fig. S6). Therefore, the dissimilar 3HA synthesis patterns suggest different GAO and PAO behavior in anaerobic metabolism. However, we require further biochemical experiments on FabF2 to confirm its effects on TFO71 anaerobic metabolism and also determine the benefits of preferential acyl-ACP integration into PHA.

Neighboring phaR and transcriptional regulator

Many *Rhodospirillales* genomes encode neighboring PhaR and a transcriptional regulator, including genomes of *Rhodospirillum rubrum* S1 (Rru_A0277 – A0278), *Magnetospirillum magnetotacticum* MS-1 (Mgn03009334 – 03009335), and *Caenispirillum salinarum* (AK4 C882_3080 – 3081).

APPENDIX B

SUPPLEMENTAL FIGURES

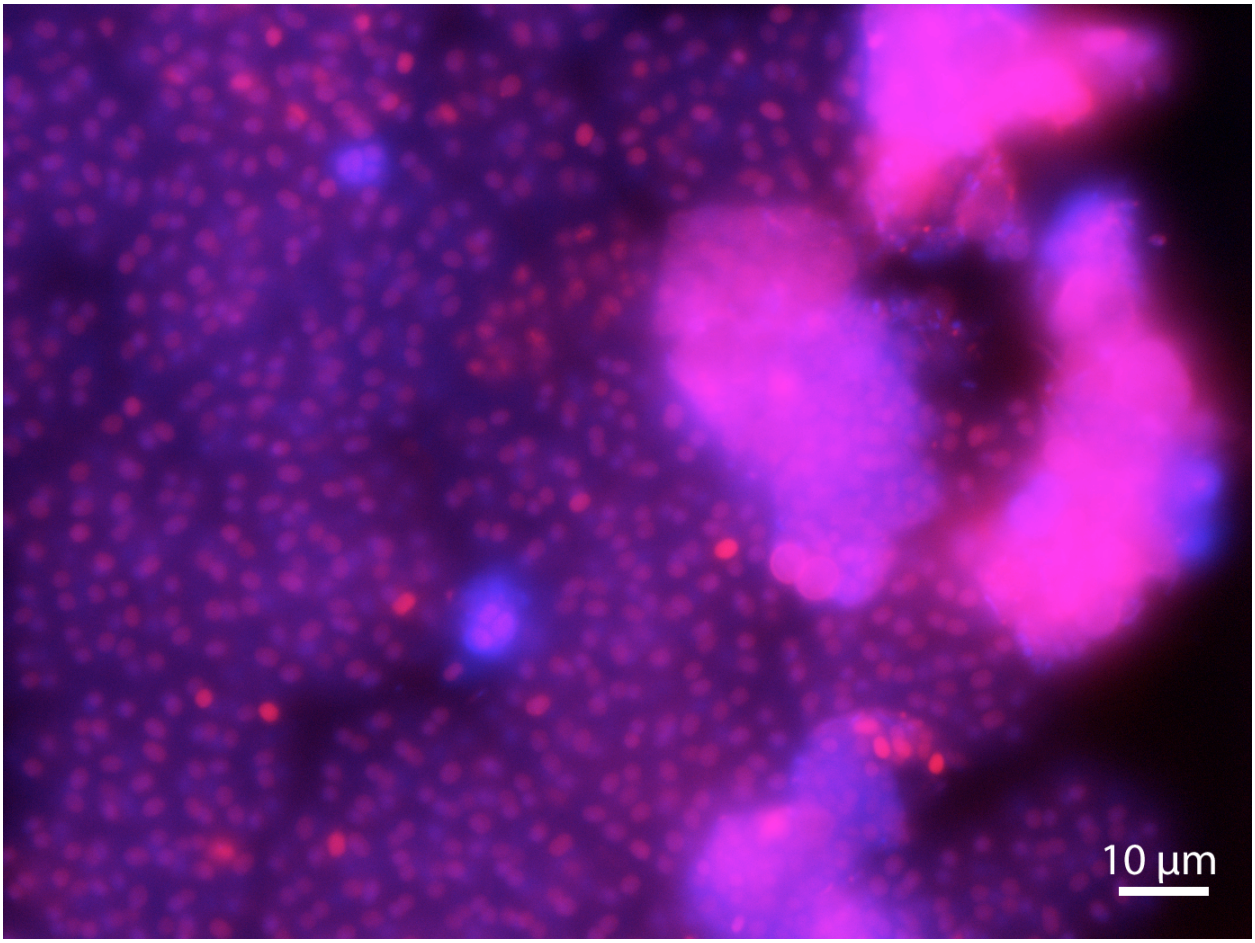


Figure S1. Fluorescence *in situ* hybridization photo of reactor DTFO community. The cyanine-3-labeled probe targets cluster I DTFO (red color). The cells were also stained with 4', 6-diamidino-2-phenylindole (blue color). DTFO cells appear purple and dominate the community.

Figure S2 (part 1 of 2)

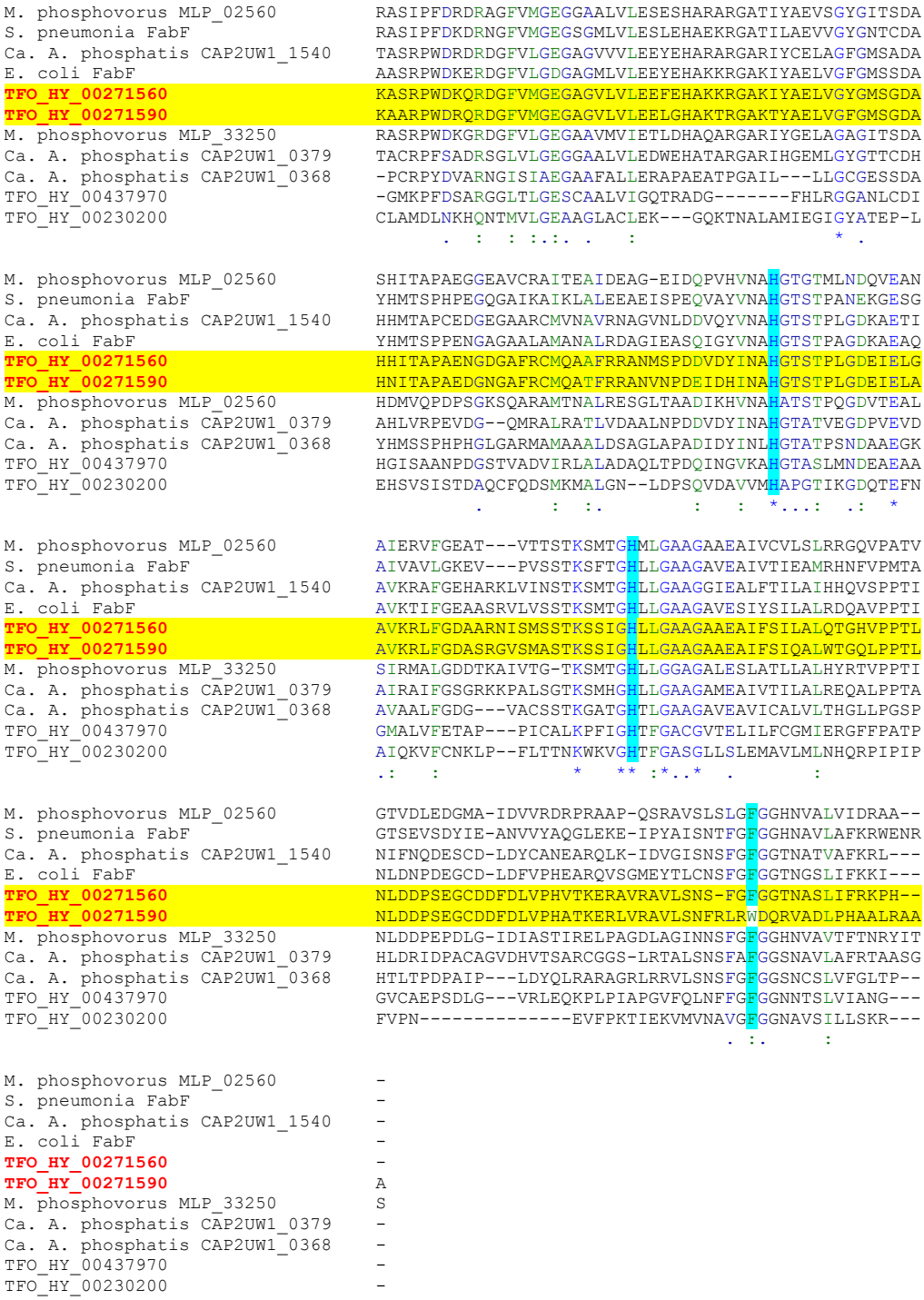


Figure S2 (part 2 of 2). Alignment of FabF from TFO71, *Candidatus* Accumulibacter phosphatis UW-1, *Microlunatus phosphovorus* NM-1, *Escherichia coli*, and *Streptococcus pneumonia* genomes. The residues essential for the acyl-ACP binding site are highlighted. Symbols: "." = weak conservation, ":" = strong conservation, "*" = full conservation. Sequences of TFO71's FabFs from the TFO_271540 – 271620 cassette are also highlighted. TFO71's TFO_HY_00271590 does not conserve the phenylalanine residue that takes part in acyl-ACP binding.

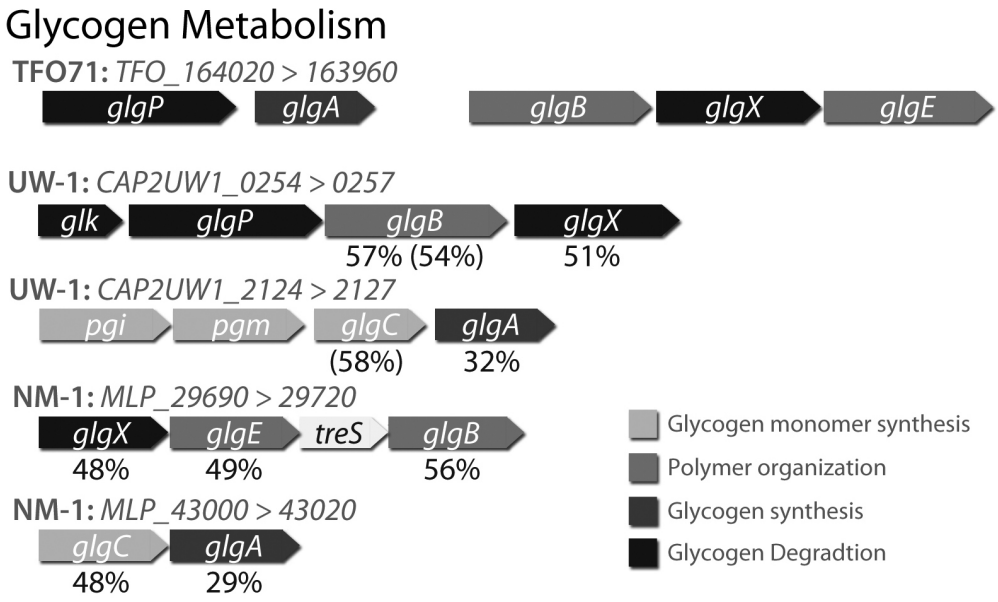


Figure S3. Comparison of representative gene cassettes between TFO71, *Candidatus* Accumulibacter phosphatis UW-1, and *Microlunatus phosphovorus* NM-1 for glycogen metabolism and synthesis. Genes are labeled with their abbreviated gene names. Refer to table S6 for exact gene product names of these gene cassettes. The coloring, as indicated in the legend, groups genes involved in same sections of glycogen metabolism and synthesis. Non-TFO71 cassettes are labeled with amino acid sequence similarity to the relevant TFO71 genes shown in the gene cassettes. Amino acid sequence similarities are labeled in parentheses when TFO71 has two copies of one gene in two different cassettes; the non-parenthesized percentage is the similarity between the non-TFO71 gene and the first corresponding TFO71 gene and the parenthesized percentage is the similarity with the second corresponding TFO71 gene.

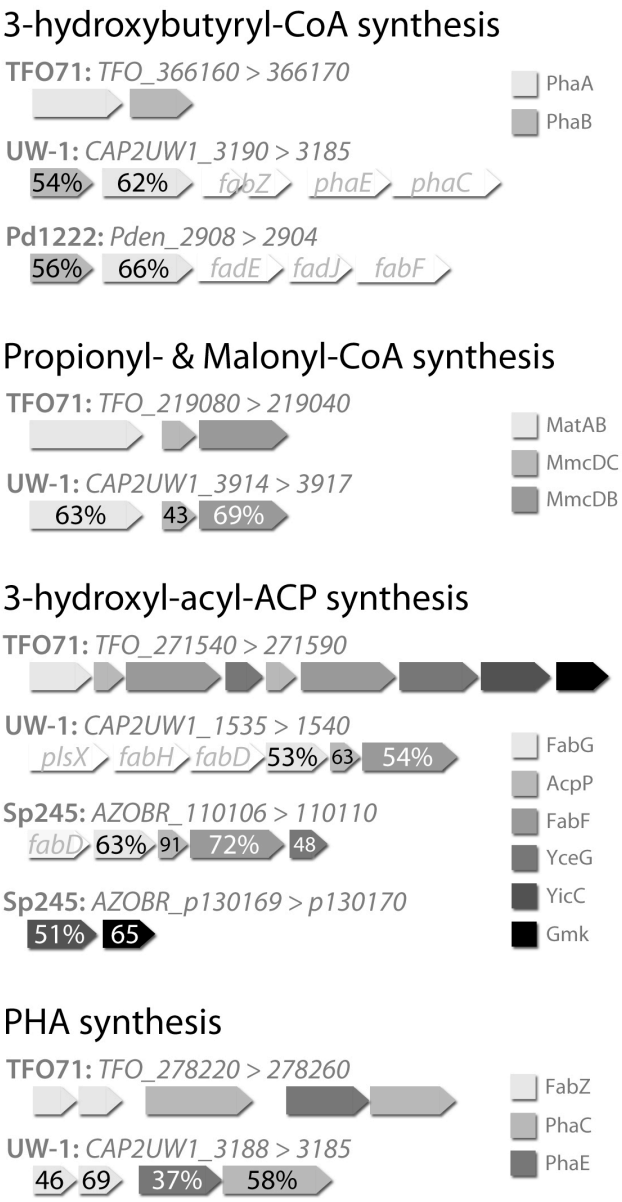


Figure S4. Comparison of representative gene cassettes between TFO71, *Candidatus* Accumulibacter phosphatis UW-1, and *Microlunatus phosphovorus* NM-1 for PHA metabolism and synthesis. Genes are labeled with their abbreviated gene names. Refer to table S7 for exact gene product names of these gene cassettes. Non-TFO71 cassettes are labeled with amino acid sequence similarity to the relevant TFO71 genes shown in the gene cassettes.

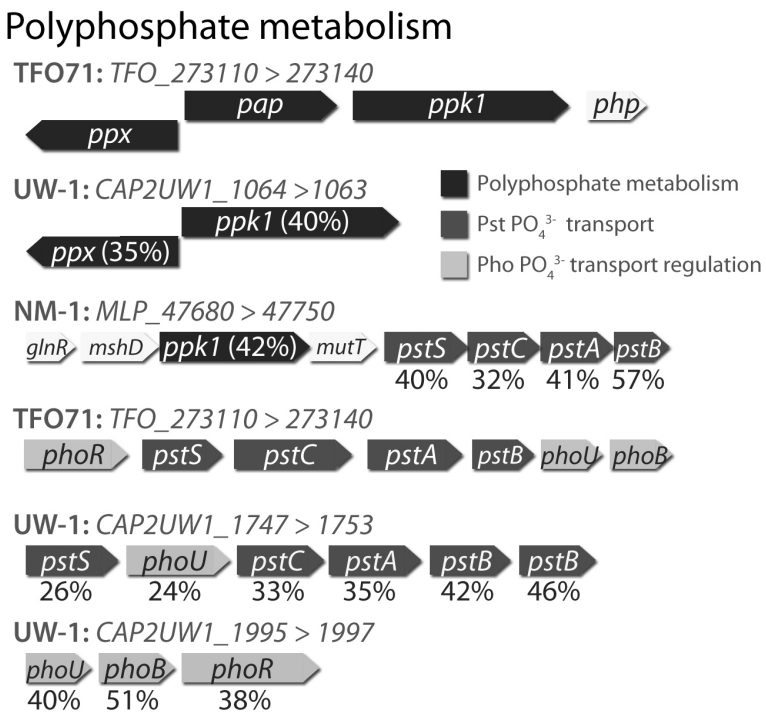
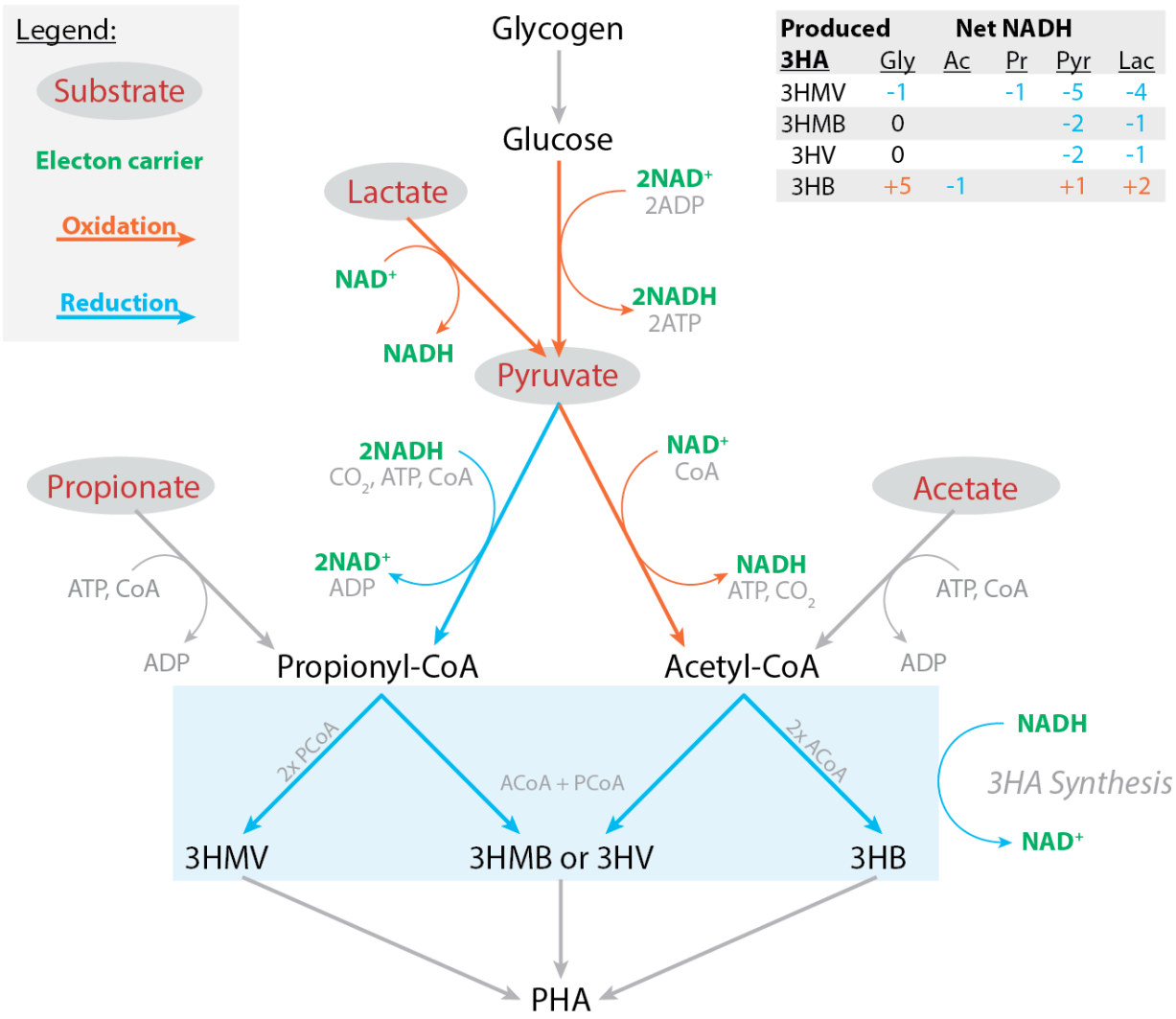


Figure S5. Comparison of representative gene cassettes between TFO71, *Candidatus* Accumulibacter phosphatis UW-1, and *Microlunatus phosphovorus* NM-1 for polyphosphate metabolism and synthesis. Genes are labeled with their abbreviated gene names. Refer to table S8 for exact gene product names of these gene cassettes. Non-TFO71 cassettes are labeled with amino acid sequence similarity to the relevant TFO71 genes shown in the gene cassettes.



Produced Net NADH					
3HA	Gly	Ac	Pr	Pyr	Lac
3HMV	-1		-1	-5	-4
3HMB	0			-2	-1
3HV	0			-2	-1
3HB	+5	-1		+1	+2

Figure S6. Simplified metabolism for PHA synthesis from glycogen and environmentally available substrates. The pathway's oxidative (orange arrows) and reductive (blue arrows) steps are shown with NAD/NADH (green) necessary for each reaction. The net NADH yield (+) or requirement (–) of 3HA synthesis from each substrate (red text with gray circle) is shown as a table in the top right. Glycolysis, lactate oxidation, and pyruvate oxidation are the primary oxidative steps, while propionyl-CoA and 3HA synthesis are reductive. It is clear that, regardless of substrate, multiple combinations of 3HA synthesis can theoretically achieve redox balance.

APPENDIX C

SUPPLEMENTAL TABLES

Table S1. General metabolism genes in TFO71's draft genome with top BLAST hits.

TCA Cycle	Abbr.	Gene product	Top BLAST hit (by amino acid sequence)	locus	Similarity (%)
TFO_HY_00357000	GltA	Citrate synthase	*Rhodospirillum rubrum* S1	Rru_A1600	76
TFO_HY_00360630	AcnA	Aconitate hydratase 1	*Azospirillum brasilense* Sp245	AZOBR_40067	73
TFO_HY_00164710	Idh2	Isocitrate dehydrogenase	*Nisaea* sp BAL199	BAL199_01824	81
TFO_HY_00354650	KorA	2-oxoglutarate ferredoxin oxidoreductase	*Kordiimonas gwangyangensis* DSM 19435	B150DRAFT_03666	63
TFO_HY_00196920	SucD	Succinyl-CoA synthetase	*Burkholderia* sp. YI23	BYI23_B010430	83
TFO_HY_00277180	SdhA	Succinate dehydrogenase flavoprotein subunit	*Magnetococcus* sp. MC-1	Magn03008820	82
TFO_HY_00277170	SdhB	Succinate dehydrogenase iron-sulfur subunit	*Nisaea* sp BAL199	BAL199_10552	76
TFO_HY_00277200	SdhC	Succinate dehydrogenase cytochrome b556 subunit	*Nisaea* sp BAL199	BAL199_10567	60
TFO_HY_00277190	SdhD	Succinate dehydrogenase membrane anchor subunit	*Azospirillum brasilense* Sp245	AZOBR_200136	57
TFO_HY_00166000	FumAB	Fumarate hydratase	*Magnetospirillum magnetotacticum* MS-1	amb1331	76
TFO_HY_00270040	Mdh	Malate dehydrogenase	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_1299	79
Glycolysis					
TFO_HY_00356720	Glk	Glucokinase	*Azospirillum lipoferum* 4B	AZOLI_0164	56
TFO_HY_00275590	Pgi	Glucose-6-phosphate isomerase	*Magnetococcus* sp. MC-1	Mmc1_1935	72
TFO_HY_00368220	Pfk	6-phosphofructokinase	*Azospirillum brasilense* Sp245	AZOBR_100017	58
TFO_HY_00358040	FdaB	Fructose-bisphosphate aldolase	*Thioalkalivibrio thiocyanodenitrificans* ARhD 1	ThithiDRAFT_1438	61
TFO_HY_00275190	TpiA	Triosephosphate isomerase	*Magnetococcus* sp. MC-1	Magn03009744	72
TFO_HY_00273980	GapA	Glyceraldehyde 3-phosphate dehydrogenase	*Pseudovibrio* sp. JE062	PJE062_2190	69
TFO_HY_00359340	Pgk	Phosphoglycerate kinase	*Fulvimarina pelagi* HTCC2506	FP2506_09186	63
TFO_HY_00267760	GpmA	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	*Lamprocystis purpurea* DSM 4197	PTSG_06592.6	67
TFO_HY_00275150	Eno	Phosphopyruvate hydratase	*Starkeya novella* DSM 506	Snov_1800	81
TFO_HY_00276810	Pyk	Pyruvate kinase	*Chamaesiphon minutus* PCC 6605	Cha6605_1128	52
Pyruvate metabolism					
TFO_HY_00354820	AceE	Pyruvate:ferredoxin oxidoreductase alpha subunit	Azospirillum brasilense Sp245	AZOBR_10525	61
TFO_HY_00354830	AceF	Pyruvate:ferredoxin oxidoreductase beta subunit	Thauera aminoaromatica MZ1T	Tmz1t_2956	64
Anaplerotic					
TFO_HY_00272280	Fbp	Fructose-1,6-bisphosphatase	*Magnetospirillum magnetotacticum* MS-1	Magn03008754	71
TFO_HY_00270050	PckA	Phosphoenolpyruvate carboxykinase (GTP)	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_1298	84
Pentose Phosphate pathway					
	Zwf	Glucose-6-phosphate 1-dehydrogenase			
	Pgl	6-phosphogluconolactonase			
TFO_HY_00358030	TktAB	Transketolase	*Kiloniella laminariae* DSM 19542	B152DRAFT_03318	60
TFO_HY_00358050	Rpe	Ribulose-phosphate 3-epimerase	*Magnetospirillum magnetotacticum* MS-1	Magn03008366	72
TFO_HY_00358040	FdaB	Fructose-bisphosphate aldolase	*Thioalkalivibrio thiocyanodenitrificans* ARhD 1	ThithiDRAFT_1438	61
TFO_HY_00364240	RpiB	Ribose 5-phosphate isomerase A	Fodinicurvata sediminis DSM 21159	G502DRAFT_2978	57
	TalAB	Fructose-6-phosphate aldolase			
Ethylmalonyl-CoA pathway					
TFO_HY_00366160	AtoB	Acetyl-CoA acetyltransferases	*Caenispirillum salinarum* AK4	C882_3458	68
TFO_HY_00267950	PhbB	Acetoacetyl-CoA reductase	*Bradyrhizobium elkanii* WSM2783	YY7DRAFT_09321	61
TFO_HY_00164310	EchA	Enoyl-CoA hydratase	Inquilinus limosus DSM 16000	G537DRAFT_00651	52
TFO_HY_00360380	Ccr	Crotonyl-CoA reductase	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_2501	71
TFO_HY_00364830	Epi	Ethylmalonyl-CoA epimerase	*Citreicella* sp. SE45	CSE45_3103	84
TFO_HY_00360390	Ecm	Ethylmalonyl-CoA mutase	*Caenispirillum salinarum* AK4	C882_2060	75
TFO_HY_00360370	Mcd	Acyl-CoA dehydrogenase	*Magnetospirillum magnetotacticum* MS-1	Magn03009555	72
TFO_HY_00268490	Mch	Mesaconyl-CoA hydratase	*Roseiflexus* sp. RS-1	RoseRS_4172	72
TFO_HY_00266220	PccA	Acetyl/propionyl-CoA carboxylase, alpha subunit	*Caenispirillum salinarum* AK4	C882_0579	60
TFO_HY_00282880	Mut	Methylmalonyl-CoA mutase, N-terminal domain/subunit	*Caenispirillum salinarum* AK4	C882_0412	57
TFO_HY_00363340	AceB	Malate synthase	*Fischerella* sp. PCC 9605	FIS9605DRAFT_01837	63

Table S2. Fermentation genes in TFO71's draft genome with top BLAST hits.

Lactate	Abbr.	Gene product	Top BLAST hit (by amino acid sequence)	locus	Similarity (%)
TFO_HY_00265840	Ldh	L-lactate dehydrogenase	*Marinithermus hydrothermalis* T1	Marky_0239	65
Ethanol					
TFO_HY_00273570	Acs	Acyl-CoA synthetase	*Roseiflexus castenholzii* HLO8	Rcas_0414	57
TFO_HY_00265060	Aldh	Aldehyde dehydrogenase (NAD+)	*Azospirillum* sp. B510	AZL_d02870	79
TFO_HY_00368870	AdhP	Alcohol dehydrogenase	*Thioflavicoccus mobilis* 8321	Thimo_1384	74
Propionate (methylmalonyl-CoA pathway)					
TFO_HY_00282880	Mut	Methylmalonyl-CoA mutase	*Rhodospirillum rubrum* S1	Rru_A2480	56
TFO_HY_00364830	MceE	Methylmalonyl-CoA epimerase	*Citreicella* sp. SE45	CSE45_3103	84
TFO_HY_00165430	MmcDC	Methylmalonyl-CoA decarboxylase	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_2101	68
TFO_HY_00219040	MmcB	Na+-transporting methylmalonyl-CoA	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_3917	67
TFO_HY_00219060	PccB	Acetyl/Propionyl-CoA carboxylase	*Candidatus Accumulibacter phosphatis* Type IIA UW-1	CAP2UW1_3914	63
TFO_HY_00266220	PccA	Acetyl/Propionyl-CoA carboxylase	*Azospirillum brasilense* Sp245	AZOBR_p1140104	60
TFO_HY_00273570	Acs	Acyl-CoA synthetase	*Roseiflexus castenholzii* HLO8	Rcas_0414	57
Succinate					
(see TCA cycle genes)					

Table S3. Respiration genes in TFO71's draft genome with top BLAST hits.

Aerobic respiration	Abbr.	Gene product	Top BLAST hit (by amino acid sequence)	locus	Similarity (%)
TFO_HY_00361570	CcoN	Cytochrome c oxidase, cbb3-type, subunit 1	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_3360	80
TFO_HY_00361560	CcoO	Cytochrome c oxidase cbb3-type subunit II	*Rhodocista centenaria* SW	RC1_1279	72
TFO_HY_00361550	CcoQ	Cytochrome c oxidase cbb3-type subunit IV	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_3358	61
TFO_HY_00361540	CcoP	Cytochrome c oxidase cbb3-type subunit III	*Agrobacterium vitis* S4	Avi_2096	56
TFO_HY_00355410	CydA	Cytochrome d ubiquinol oxidase subunit I	Leisingera nanhaiensis NH52F	Leina_03651	61
TFO_HY_00355420	CydB	Cytochrome d ubiquinol oxidase subunit II	Rhizobium sp. CF122	PMI09_04890	59
F-type ATPase					
TFO_HY_00205910	AtpH	F-type H+-transporting ATPase subunit delta	*Rhodospirillum photometricum* DSM 122	RSPPHO_01464	52
TFO_HY_00205920	AtpA	F-type H+-transporting ATPase subunit alpha	*Rhodocista centenaria* SW	RC1_2243	79
TFO_HY_00205930	AtpG	F-type H+-transporting ATPase subunit gamma	*Rhodocista centenaria* SW	RC1_2242	62
TFO_HY_00205940	AtpD	F-type H+-transporting ATPase subunit beta	*Azospirillum brasilense* Sp245	AZOBR_40306	83
TFO_HY_00205950	AtpC	F-type H+-transporting ATPase subunit epsilon	*Tistrella mobilis* KA081020-065	TMO_2984	46
TFO_HY_00264970	AtpB	F-type H+-transporting ATPase subunit a	*Nisaea* sp BAL199	BAL199_08868	63
TFO_HY_00264980	AtpE	F-type H+-transporting ATPase subunit c	*Tistrella mobilis* KA081020-065	TMO_2780	64
TFO_HY_00264990	AtpF	F-type H+-transporting ATPase subunit b	*Magnetospirillum magnetotacticum* MS-1	Magn03010598	60
Nitrate respiration					
TFO_HY_00264480	NrtA	Nitrate ABC transporter, periplasmic	*Microvirga* sp. WSM3557	MicloDRAFT_00032520	47
TFO_HY_00264490	NrtA	Nitrate ABC transporter, periplasmic	*Magnetospirillum magneticum* AMB-1	amb0537	80
TFO_HY_00264500	NrtB	Nitrate ABC transporter, permease	*Azospirillum brasilense* Sp245	AZOBR_p1100073	72
TFO_HY_00264510	NrtD	Nitrate ABC transporter, ATPase	*Magnetospirillum magneticum* AMB-1	amb0535	72
TFO_HY_00264530	NirB	Ferredoxin-nitrite reductase	*Rhodopseudomonas palustris* BisA53	RPE_1261	60
TFO_HY_00264540	SiR	Sulfite reductase, α subunit (flavoprotein)	Bradyrhizobium sp. ORS 278	BRADO3794	58
TFO_HY_00264550	NasA	Nitrate reductase / MPT oxidoreductase	*Rhodopseudomonas palustris* BisA53	RPB_1752	59
TFO_HY_00264560		Nitroreductase	*Burkholderiales* sp. JOSHI_001	BurJ1DRAFT_3523	42
TFO_HY_00264570	MoeA	Molybdopterin biosynthesis enzyme	*Methylobacterium extorquens* AM1	MexAM1_META2p0368	46
TFO_HY_00264580	MobA	Molybdenum cofactor guanylyltransferase	*Rhodopseudomonas palustris* BisA53	Rpal_4175	43
TFO_HY_00180870	Nir	Nitrite reductase, copper-containing	Hyphomicrobium denitrificans 1NES1	Hypde_03101	66
TFO_HY_00180720	NorB	Nitric oxide reductase large subunit	*Rhodocista centenaria* SW	RC1_3976	65
TFO_HY_00364660	NuoA	NADH:ubiquinone oxidoreductase subunit 3	*Tistrella mobilis* KA081020-065	TMO_1978	75
TFO_HY_00364670	NuoB	NADH-ubiquinone oxidoreductase chain B	*Caenispirillum salinarum* AK4	C882_4191	80
TFO_HY_00364680	NuoC	NADH-ubiquinone oxidoreductase chain C	*Nisaea* sp BAL199	BAL199_25409	67
TFO_HY_00364690	NuoD	NADH dehydrogenase D subunit	*Micavibrio aeruginosavorus* ARL-13	MICA_1368	75
TFO_HY_00364700	NuoE	NADH dehydrogenase I chain E	*Azospirillum* sp. B510	AZL_021440	64
TFO_HY_00364710	NuoF	NADH-quinone oxidoreductase, F subunit	*Parvibaculum lavamentivorans* DS-1	Plav_3221	77
TFO_HY_00364720	NuoG	NADH dehydrogenase, G subunit	*Azospirillum brasilense* Sp245	AZOBR_100249	65
TFO_HY_00364730	NuoH	NADH dehydrogenase subunit H	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_2711	75
TFO_HY_00364740	NuoI	NADH-quinone oxidoreductase, chain I	*Kiloniella laminariae* DSM 19542	B152DRAFT_02585	86
TFO_HY_00364750	NuoJ	NADH dehydrogenase subunit J	*Nisaea* sp BAL199	BAL199_25444	71
TFO_HY_00364760	NuoK	NADH:ubiquinone oxidoreductase subunit K	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_2708	86
TFO_HY_00364770	NuoL	NADH-ubiquinone oxidoreductase chain L	*Magnetospirillum magnetotacticum* MS-1	Magn03007561	69
TFO_HY_00364780	NuoM	NADH-quinone oxidoreductase, chain M	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_2706	73
TFO_HY_00364790	NuoN	NADH dehydrogenase subunit N	*Azospirillum brasilense* Sp245	AZOBR_100256	61
TFO_HY_00219380	PetA	Ubiquinol-cytochrome c reductase, iron-sulfur subunit	*Caenispirillum salinarum* AK4	C882_2242	71
TFO_HY_00219370	PetB	Cytochrome b subunit of the bc complex	*Magnetospirillum magneticum* AMB-1	amb4089	72
TFO_HY_00357950	NirA	Nitrite/Sulphite reductase	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_0407	72

Table S4. Nitrogen fixation genes in TFO71's draft genome with top BLAST hits.

Nitrogen fixation	Abbr.	Gene product	Top BLAST hit (by amino acid sequence)	locus	Similarity (%)
TFO_HY_00357650	NifH	Nitrogenase iron protein	*Rhizobium mongolense* USDA 1844	A3C3DRAFT_06795	83
TFO_HY_00357640	NifD	Nitrogenase Mo-Fe protein alpha chain	*Methylocystis* sp. Rockwell	Met49242DRAFT_2939	77
TFO_HY_00357630	NifK	Nitrogenase Mo-Fe protein beta chain	*Rhodobacter capsulatus* SB1003	RCAP_rcc00570	68
TFO_HY_00357620	NifE	Nitrogenase Mo-Fe protein alpha/beta chain	*Methylosinus* sp. LW3	METLW3DRAFT_3345	64
TFO_HY_00357610	NifN	Nitrogenase Mo-Fe cofactor biosynthesis protein	*Caenispirillum salinarum* AK4	C882_4064	58
TFO_HY_00357600	NifX	Nitrogen fixation protein	*Caenispirillum salinarum* AK4	C882_4063	58
TFO_HY_00357590		Nitrogen fixation protein	*Mesorhizobium ciceri bv biserrulae* WSM1271	Mesci_5811	55
TFO_HY_00357580		4Fe-4S ferredoxin, nitrogenase-associated	*Caenispirillum salinarum* AK4	C882_4060	55
TFO_HY_00357570	NifQ	Nitrogenase Mo delivery protein	*Caenispirillum salinarum* AK4	C882_4059	61
TFO_HY_00357560	IscA1	Fe-S cluster assembly protein	*Methylobacterium* sp. 4-46	M446_3583	51
TFO_HY_00357550	NifV	Homocitrate synthase	*Caenispirillum salinarum* AK4	C882_3793	63
TFO_HY_00357540	NifW	Nitrogenase-stabilizing protein	*Rhodomicrobium vannielii* ATCC 17100	Rvan_1126	45
TFO_HY_00358290	FixA	Electron transfer flavoprotein FixA	*Kangiella aquimarina* DSM 16071	B158DRAFT_1452	69
TFO_HY_00358300	FixB	Electron transfer flavoprotein FixB	*Azospirillum brasilense* Sp245	AZOBR_p140082	73
TFO_HY_00265200	FixC	Electron transfer flavoprotein FixC	*Desulfotomaculum acetoxidans* 5575	Dtox_4276	52

Table S5. Biosynthesis genes in TFO71's draft genome with top BLAST hits.

Aminoacyl tRNA synthetases			Top BLAST hit	Locus	Similarity (%)
		Alanyl-tRNA synthetase			
TFO_HY_00158680		Arginyl-tRNA synthetase	*Rhodospirillum photometricum* DSM 122	RSPPHO_03008	63
TFO_HY_00273080		Aspartyl-tRNA synthetase	*Rhodospirillum rubrum* S1	Rru_A2072	73
TFO_HY_00178820		Cysteinyl-tRNA synthetase	*Azospirillum brasilense* Sp245	AZOBR_40357	66
TFO_HY_00357010		Glutamyl- and glutaminyl-tRNA synthetases	*Rhodospirillum rubrum* S1	Rru_A1601	64
TFO_HY_00359990		Glycyl-tRNA synthetase, alpha subunit	*Magnetospirillum magnetotacticum* MS-1	Magn03008287	83
TFO_HY_00197040		Glycyl-tRNA synthetase, beta subunit	*Azospirillum brasilense* Sp245	AZOBR_180064	64
TFO_HY_00219920		Histidyl-tRNA synthetase	*Rhodospirillum rubrum* S1	Rru_A0748	64
TFO_HY_00219420		Isoleucyl-tRNA synthetase	*Magnetospirillum magnetotacticum* MS-1	Magn03010800	66
TFO_HY_00368130		Leucyl-tRNA synthetase	*Azospirillum brasilense* Sp245	AZOBR_10408	66
TFO_HY_00358180		Methionyl-tRNA synthetase	*Rhodospirillum rubrum* S1	Rru_A1701	75
TFO_HY_00269490		Phenylalanyl-tRNA synthetase alpha subunit	*Rhodospirillum rubrum* S1	Rru_A3523	68
TFO_HY_00269500		Phenylalanyl-tRNA synthetase beta subunit	*Azospirillum* sp. B510	AZL_025980	63
TFO_HY_00364870		Prolyl-tRNA synthetase	*Azospirillum brasilense* Sp245	AZOBR_100265	71
TFO_HY_00158770		Seryl-tRNA synthetase	*Magnetospirillum magneticum* AMB-1	amb2521	69
TFO_HY_00275830		Threonyl-tRNA synthetase	*Rhodospirillum rubrum* S1	Rru_A3487	78
TFO_HY_00366880		Tryptophanyl-tRNA synthetase	*Rhodospirillum photometricum* DSM 122	RSPPHO_01504	71
		Tyrosyl-tRNA synthetase			
TFO_HY_00164380		Valyl-tRNA synthetase	*Azospirillum lipoferum 4B*	AZOLI_1559	68
Aspartate and Asparagine					
TFO_HY_00220140	AspB	Apartate/tyrosine/aromatic amino transferase	*Azospirillum* sp. B510	AZL_001910	74
TFO_HY_00178750	AsnB	Asparagine synthase	*Labrenzia aggregata* IAM 12614	SIAM614_16722	74
TFO_HY_00270320	GudB	NAD-specific glutamate dehydrogenase	*Caenispirillum salinarum* AK4	C882_3179	52
TFO_HY_00357990	GlnA	Glutamate synthetase, type I	*Caenispirillum salinarum* AK4	C882_1090	78
Glutamate and Glutamine					
TFO_HY_00267030	ProB	Glutamate kinase	*Magnetospirillum magneticum* AMB-1	amb4080	62
TFO_HY_00267040	ProA	Glutamyl phosphate reductase	*Caenispirillum salinarum* AK4	C882_2494	67
TFO_HY_00276370	ProC	Pyrroline-5-carboxylate reductase	*Azospirillum brasilense* Sp245	AZOBR_100037	59
Arginine and Proline					
	Ocd	Ornithine cyclodeaminase			
TFO_HY_00268240	RocD	Ornithine/acetylornithine aminotransferase	*Methylococcus capsulatus* Bath	MCA0809	64
TFO_HY_00269910	Otc	Ornithine carbamoyltransferase	*Bradyrhizobium* sp. SA-4 CB756	BrageDRAFT_7234	64
TFO_HY_00358950	CarA	Carbamoyl-phosphate synthase, small subunit	*Azospirillum* sp. B510	AZL_010540	74
TFO_HY_00358940	CarB	Carbamoyl-phosphate synthase, large subunit	*Magnetospirillum magneticum* AMB-1	amb0702	73
TFO_HY_00363700	ArgG	Argininosuccinate synthase	*Tistrella mobilis* KA081020-065	TMO_3105	81
TFO_HY_00279380	ArgH	Argininosuccinate lyase	*Rhodocista centenaria* SW	RC1_4103	66
Cysteine and Methionine					
TFO_HY_00196810	CysE	Serine O-acetyltransferase	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_2774	57
TFO_HY_00220460	CysK	Cysteine synthase	*Rhodospirillum photometricum* DSM 122	RSPPHO_03289	76
TFO_HY_00273520	MetC	Cystathionine beta-lyase	*Azospirillum* sp. B510	AZL_012890	57
TFO_HY_00220360	MetH	Methionine synthase I	*Kordiimonas gwangyangensis* DSM 19435	B150DRAFT_02904	66
Glycine, Serine, and Threonine					
TFO_HY_00159440		Threonine synthase	*Inquilinus limosus DSM 16000*	G537DRAFT_02230	65
TFO_HY_00364250		Glycine/serine hydroxymethyltransferase	*Inquilinus limosus DSM 16000*	G537DRAFT_03261	75
TFO_HY_00270150		Glycine cleavage system T protein	*Azospirillum halopraeferens DSM 3675*	G472DRAFT_03536	58
TFO_HY_00270140		Glycine cleavage system H protein	*Inquilinus limosus DSM 16000*	G537DRAFT_00886	63
		Glycine cleavage system protein P (pyridoxal-			
TFO_HY_00270130		binding), N-terminal domain	*Azospirillum sp. B510*	AZL_015860	70
TFO_HY_00364250		Glycine/serine hydroxymethyltransferase	*Methylocaldum szegediense O-12*	MetszDRAFT_2018	70
Histidine					
TFO_HY_00270780	HisG	ATP phosphoribosyltransferase	*Caenispirillum salinarum* AK4	C882_1783	72
TFO_HY_00164900	HisE	Phosphoribosyl-ATP pyrophosphohydrolase	*Caenispirillum salinarum* AK4	C882_2996	62
TFO_HY_00164160	HisI	Phosphoribosyl-AMP cyclohydrolase	*Roseibium* sp. TrichSKD4	TRICHSKD4_6127	75
		Phosphoribosylformimino-5-aminoimidazole			
TFO_HY_00279820	HisA	carboxamide ribotide isomerase	*Magnetospirillum magneticum* AMB-1	amb4532	65
		Imidazoleglycerol phosphate synthase, cyclase			
TFO_HY_00164890	HisF	subunit	*Rhodospirillum photometricum* DSM 122	RSPPHO_01063	87
		Imidazole glycerol phosphate synthase, glutamine			
TFO_HY_00279830	HisH	amidotransferase subunit	*Rhodocista centenaria* SW	RC1_3296	71
TFO_HY_00279850	HisB	Imidazoleglycerol-phosphate dehydratase	*Sphingomonas melonis* DAPP-PG 224	Sphme3DRAFT_1925	73
TFO_HY_00271180	HisC	Histidinol-phosphate transaminase	*Azospirillum brasilense* Sp245	AZOBR_p1130163	62
TFO_HY_00270790	HisD	Histidinol dehydrogenase	*Magnetospirillum magneticum* AMB-1	amb3342	69
Lysine					
TFO_HY_00271080	LysC	Aspartate kinase	*Magnetospirillum magneticum* AMB-1	amb1612	75
TFO_HY_00276580	Asd	Aspartate-semialdehyde dehydrogenase	*Rhodocista centenaria* SW	RC1_3933	76
TFO_HY_00157950	DapA	Dihydrodipicolinate synthase	*Methylobacterium nodulans* ORS 2060	Mnod_6692	70
TFO_HY_00275270	DapB	Dihydrodipicolinate reductase	*Rhodocista centenaria* SW	RC1_3162	64
TFO_HY_00158370	DapD	Tetrahydrodipicolinate N-succinyltransferase	*Methylobacterium extorquens* AM1	MexAM1_META1p1465	71
TFO_HY_00269900	ArgD	Succinyldiaminopimelate transaminase	*Tistrella mobilis* KA081020-065	TMO_0027	62
TFO_HY_00158380	DapE	Succinyl-diaminopimelate desuccinylase	*Magnetospirillum magneticum* AMB-1	amb3874	63
TFO_HY_00264210	DapF	Diaminopimelate epimerase	*Azospirillum brasilense* Sp245	AZOBR_200186	61
TFO_HY_00217790	LysA	Diaminopimelate decarboxylase	*Magnetospirillum magnetotacticum* MS-1	Magn03010258	61

Phenylalanine, Tyrosine, and Tryptophan					
TFO_HY_00178780	AroFGH	3-deoxy-7-phosphoheptulonate synthase	*Azospirillum lipoferum* 4B	AZOLI_p10482	70
TFO_HY_00275230	AroB	3-dehydroquinate synthase	*Caenispirillum salinarum* AK4	C882_2124	61
TFO_HY_00167430	AroQ	3-dehydroquinate dehydratase II	*Rhodomicrobium vannielii* ATCC 17100	Rvan_3422	65
TFO_HY_00367100	AroE	Shikimate 5-dehydrogenase	*Azospirillum brasilense* Sp245	AZOBR_10392	59
TFO_HY_00275240	AroKL	Shikimate kinase	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_3222	60
TFO_HY_00361280	AroA	3-phosphoshikimate 1-carboxyvinyltransferase	*Magnetospirillum magnetotacticum* MS-1	Magn03010163	65
TFO_HY_00266400	AroC	Chorismate synthase	*Magnetospirillum magneticum* AMB-1	amb1590	72
TFO_HY_00276960	TrpE	Anthranilate synthase	*Rhodocista centenaria* SW	RC1_0109	62
TFO_HY_00365040	TrpD	Anthranilate phosphoribosyltransferase	*Rhodospirillum photometricum* DSM 122	RSPPHO_00543	67
TFO_HY_00361080	TrpF	Phosphoribosylanthranilate isomerase	*Silicibacter lacuscaerulensis* ITI-1157	SL1157_0262	54
TFO_HY_00365050	TrpC	Indole-3-glycerol phosphate synthase	*Rhodocista centenaria* SW	RC1_0103	60
TFO_HY_00361100	TrpA	Tryptophan synthase alpha subunit	*Rhodocista centenaria* SW	RC1_2830	67
TFO_HY_00361090	TrpB	Tryptophan synthase beta subunit	*Parvibaculum lavamentivorans* DS-1	Plav_0143	79
TFO_HY_00355440	PheA2	Prephenate dehydratase	*Rhodocista centenaria* SW	RC1_3406	65
TFO_HY_00220140	AspB	Apartate/tyrosine/aromatic amino transferase	*Azospirillum* sp. B510	AZL_001910	74
TFO_HY_00271170	TyrA2	Prephenate dehydrogenase	*Rhodocista centenaria* SW	RC1_4086	66
Isoleucine, Leucine, and Valine					
TFO_HY_00356110	IlvBGI	Acetolactate synthase			
TFO_HY_00263840	IlvC	Ketol-acid reductoisomerase	*Rhodospirillum rubrum* S1	Rru_A0469	80
TFO_HY_00158560	IlvD	Dihydroxy-acid dehydratase	*Rhodocista centenaria* SW	RC1_1163	81
TFO_HY_00276280	IlvE	Branched-chain-amino-acid transaminase	*Rhodospirillum rubrum* S1	Rru_A2223	69
TFO_HY_00219700	LeuA	2-isopropylmalate synthase	*Rhodospirillum rubrum* S1	Rru_A0695	73
TFO_HY_00276540	LeuD	3-isopropylmalate dehydratase small subunit	*Oceanicola granulosus* HTCC2516	OG2516_09580	68
TFO_HY_00276530	LeuC	3-isopropylmalate dehydratase large subunit	*Caenispirillum salinarum* AK4	C882_4363	78
TFO_HY_00276550	LeuB	3-isopropylmalate dehydrogenase	*Meganema perideroedes* DSM 15528	B161DRAFT_01930	75
Ribosomal protein					
TFO_HY_00252940	L1	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_2940	75
TFO_HY_00360710	L2	Ribosomal protein	*Magnetospirillum magnetotacticum* MS-1	Magn03008532	81
TFO_HY_00360680	L3	Ribosomal protein	*Nisaea* sp BAL199	BAL199_29755	70
TFO_HY_00360690	L4	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_0615	64
TFO_HY_00360800	L5	Ribosomal protein	*Rhodocista centenaria* SW	RC1_0723	70
TFO_HY_00360830	L6	Ribosomal protein	*Rhodocista centenaria* SW	RC1_0726	61
TFO_HY_00271510	L9	Ribosomal protein	*Acetobacter pasteurianus* IFO 3283-12	APA12_09950	62
TFO_HY_00252930	L10	Ribosomal protein	*Azospirillum lipoferum* 4B	AZOLI_0448	65
TFO_HY_00252950	L11	Ribosomal protein	*Rhodospirillum photometricum* DSM 122	RSPPHO_00017	66
TFO_HY_00357890	L13	Ribosomal protein	*Candidatus Puniceispirillum marinum* IMCC1322	SAR116_0493	71
TFO_HY_00360780	L14	Ribosomal protein	*Parvibaculum lavamentivorans* DS-1	Plav_2746	85
TFO_HY_00360870	L15	Ribosomal protein	*Geminicoccus roseus* DSM 18922	GemroDRAFT_2893	64
TFO_HY_00360750	L16	Ribosomal protein	*Labrenzia aggregata* IAM 12614	SIAM614_16157	82
TFO_HY_00360930	L17	Ribosomal protein	*Ahrensia* sp. R2A130	R2A130_0137	64
TFO_HY_00360840	L18	Ribosomal protein	*Rhodospirillum rubrum* S1	Rru_A2672	64
TFO_HY_00276520	L19	Ribosomal protein	*Magnetospirillum magnetotacticum* MS-1	Magn03009316	71
TFO_HY_00269480	L20	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_2478	78
TFO_HY_00267000	L21	Ribosomal protein	*Candidatus Puniceispirillum marinum* IMCC1322	SAR116_1856	45
TFO_HY_00360730	L22	Ribosomal protein	*Rhodovibrio salinarum* DSM 9154	RHOSADRAFT_3095	71
TFO_HY_00360700	L23	Ribosomal protein	*Rhodospirillum rubrum* S1	Rru_A2686	70
TFO_HY_00360790	L24	Ribosomal protein	*Azospirillum* sp. B510	AZL_005920	72
TFO_HY_00267010	L27	Ribosomal protein	*Oceaniovalibus guishaninsula* JLT2003	OCGS_0149	79
TFO_HY_00360760	L29	Ribosomal protein	*Jannaschia* sp. CCS1	Jann_0599	65
TFO_HY_00360860	L30	Ribosomal protein	*Rhodospirillum rubrum* S1	Rru_A2670	73
TFO_HY_00206930	L33	Ribosomal protein	*Magnetospirillum magnetotacticum* MS-1	Magn03010711	84
TFO_HY_00269470	L35	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_2477	69
TFO_HY_00281130	L36	Ribosomal protein	*Roseomonas cervicalis* ATCC 49957	HMPREF0731_0587	78
TFO_HY_00361040	S1	Ribosomal protein	*Azospirillum brasilense* Sp245	AZOBR_40092	76
TFO_HY_00364910	S2	Ribosomal protein	*Azospirillum lipoferum* 4B	AZOLI_1070	78
TFO_HY_00360740	S3	Ribosomal protein	*Magnetospirillum magneticum* AMB-1	amb3124	79
TFO_HY_00277810	S4	Ribosomal protein	*Thalassobaculum salexigens DSM 19539*	G578DRAFT_2418	77
TFO_HY_00360850	S5	Ribosomal protein	*Rhodocista centenaria* SW	RC1_0728	68
TFO_HY_00252900	S7	Ribosomal protein	*Magnetospirillum magneticum* AMB-1	amb3134	78
TFO_HY_00360820	S8	Ribosomal protein	*Rhodocista centenaria* SW	RC1_0725	77
TFO_HY_00357900	S9	Ribosomal protein	*Commensalibacter intestini* A911	CIN_19380	68
TFO_HY_00360670	S10	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_0617	91
TFO_HY_00360910	S11	Ribosomal protein	*Magnetospirillum magneticum* AMB-1	amb3109	85
TFO_HY_00252910	S12	Ribosomal protein	*Kiloniella laminariae* DSM 19542	B152DRAFT_04120	90
TFO_HY_00360900	S13	Ribosomal protein	*Tistrella mobilis* KA081020-065	TMO_2610	85
TFO_HY_00360810	S14	Ribosomal protein	*Pseudovibrio* sp. JE062	PJE062_4677	68
TFO_HY_00277950	S15	Ribosomal protein	*Kiloniella laminariae DSM 19542*	B152DRAFT_00250	81
TFO_HY_00276490	S16	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_4359	61
TFO_HY_00360770	S17	Ribosomal protein	*Rhodocista centenaria* SW	RC1_0720	77
TFO_HY_00271530	S18	Ribosomal protein	*Labrys methylaminiphilus* JLW10	LABMEDRAFT_3866	94
TFO_HY_00360720	S19	Ribosomal protein	*Caenispirillum salinarum* AK4	C882_0612	82
TFO_HY_00220540	S20	Ribosomal protein	*Azospirillum brasilense* Sp245	AZOBR_10279	72
TFO_HY_00266530	S21	Ribosomal protein	*Magnetospirillum magneticum* AMB-1	amb2820	96

Table S6. Glycogen metabolism genes in TFO71's draft genome with top BLAST hits and top hits in genomes of *Candidatus Accumulibacter phosphatis* UW-1 and *Microlunatus phosphovorus* NM-1.

TFO71 locus	Gene product		Top BLAST hit	Locus	%ID	UW-1 locus	%ID	NM-1 locus	%ID
TFO_HY_00164020	GlgP	Glycogen phosphorylase	*Spirochaeta smaragdinae* DSM 11293	Spirs_0940	67	CAP2UW1_0255	49	MLP_46800	48
TFO_HY_00164010	GlgA	Glycogen synthase, ADP-glucose type	*Methylococcus capsulatus* str. Bath	MCA2606	66	CAP2UW1_2127	32		
TFO_HY_00163980	GlgB	Glycogen branching enzyme	*Azospirillum brasilense* Sp245	AZOBR_140107	63	CAP2UW1_0256	57	MLP_29720	49
TFO_HY_00163970	GlgX	Glycogen debranching enzyme	*Azospirillum brasilense* Sp245	AZOBR_140110	60	CAP2UW1_0257	51	MLP_36670/ MLP_29690	51/48
TFO_HY_00163960	GlgE	4-alpha-glucanotransferase	*Azospirillum brasilense* Sp245	AZOBR_140112	46			MLP_15250	35
TFO_HY_00266730	Pph	Protein phosphatase	*Azospirillum brasilense* Sp245	AZOBR_p1160006	51	CAP2UW1_4137	31	MLP_33770	27
TFO_HY_00266740	GlgA	Glycogen synthase	*Rhodovulum* sp. PH10	A33M_1631	65	CAP2UW1_2661	54	MLP_43020	47
TFO_HY_00266750	GlgX	Glycogen debranching enzyme	*Caenispirillum salinarum* AK4	C882_0985	55	CAP2UW1_2662	25	MLP_29700	24
TFO_HY_00266760	GlgE	4-alpha-glucanotransferase	*Caenispirillum salinarum* AK5	C882_0986	62	CAP2UW1_2662	49	MLP_29700	44
TFO_HY_00278210	GlgA	Glycogen synthase	*Methylacidiphilum fumariolicum* SolV	MFUM_40010	47	CAP2UW1_2661	26	MLP_42560	31
TFO_HY_00278200	GlgX	Glycogen debranching enzyme	*Methylomicrobium alcaliphilum* 20Z	MEALZ_0488	70				
TFO_HY_00272190	Pgm	Phosphoglucomutase	*Nodosilinea nodulosa* PCC 7104	Lepto7104DRAFT_2059	73	CAP2UW1_2125	65	MLP_29820	24
TFO_HY_00273590	GlgC	Glucose-1-phosphate, adenylyl transferase	*Thiorhodovibrio* sp. 970	Thi970DRAFT_0817	72	CAP2UW1_2126	42	MLP_43000	44

Table S7. Polyhydroxyalkanoate metabolism genes in TFO71's draft genome with top BLAST hits and top hits in genomes of *Candidatus Accumulibacter phosphatis* UW-1 and *Microlunatus phosphovorus* NM-1.

TFO71 locus	Gene product		Top BLAST hit	Locus	%ID	UW-1 locus	%ID	NM-1 locus	%ID
TFO_HY_00278220	PhaJ	Enoyl-CoA hydratase (MaoC-like)	*Parvibaculum lavamentivorans* DS-1	Plav_0156	59	CAP2UW1_3188	46		
TFO_HY_00278230	PhaJ	Enoyl-CoA hydratase (MaoC-like)	*C. A. phosphatis* UW-1	CAP2UW1_3187	60	CAP2UW1_3187	60		
TFO_HY_00278240	PhaC	PHA synthase, class III, PhaC subunit	*C. A. phosphatis* UW-1	CAP2UW1_3193	72	CAP2UW1_3185	24		
TFO_HY_00278250	PhaE	PHA synthase PhaE subunit	*C. A. phosphatis* UW-1	CAP2UW1_3186	37	CAP2UW1_3186	37		
TFO_HY_00278260	PhaC	PHA synthase, class III, PhaC subunit	*C. A. phosphatis* UW-1	CAP2UW1_3185	58	CAP2UW1_3185	58		
TFO_HY_00366160	PhaA	Acetyl-CoA acetyltransferase	*Acidovorax avenae* ATCC 19860	Acav_4265	66	CAP2UW1_3189	62		
TFO_HY_00366170	PhaB	Short-chain alcohol dehydrogenase	*Azospirillum brasilense* Sp245	AZOBR_p280144	60	CAP2UW1_3190	54		
TFO_HY_00271540	FabG	3-ketoacyl-(ACP) reductase	*Rhodobacter capsulatus* SB 1003	RCAP_rcc01676	63	CAP2UW1_1538	53	MLP_14720	48
TFO_HY_00271550	AcpP	Acyl carrier protein	*Azospirillum brasilense* Sp245	AZOBR_110108	91	CAP2UW1_1539	63	MLP_33240	44
TFO_HY_00271560	FabF	Beta-ketoacyl-(ACP) synthase II	*Oceanibaculum indicum* P24	P24_04140	75	CAP2UW1_1540	54	MLP_33250	36
TFO_HY_00271570	YceG	Septation-associated periplasmic protein	*Azospirillum brasilense* Sp245	AZOBR_110110	48	CAP2UW1_1629	29		
TFO_HY_00271580	AcpP	Acyl carrier protein	*Thalassospira profundimaris* WP0211	TH2_11179	62	CAP2UW1_1539	49	MLP_33240	38
TFO_HY_00271590	FabF	Beta-ketoacyl-(ACP) synthase II	*Oceanibaculum indicum* P24	P24_04140	67	CAP2UW1_1540	49	MLP_33250	30
TFO_HY_00271600	YceG	Septation-associated periplasmic protein	*Azospirillum brasilense* Sp245	AZOBR_110110	54	CAP2UW1_2472	37	MLP_26490	38
TFO_HY_00271610	YicC	Stress-induced protein	*Oceanibaculum indicum* P24	P24_04045	52	CAP2UW1_0934	37		
TFO_HY_00271620	Gmk	Guanylate kinase	*Bradyrhizobium japonicum* USDA 110	blr4088	69	CAP2UW1_0932	50	MLP_26340	44
TFO_HY_00361600	FabD	Malonyl-CoA-ACP transacylase	*Rhodocista centenaria SW*	RC1_1572	65	CAP2UW1_1537	44	MLP_33230	32
TFO_HY_00266630	PhaZ	PHA depolymerase	*Methylocystis rosea* SV97T	A3OODRAFT_1661	44	CAP2UW1_1898	43		
TFO_HY_00219210	PhaZ	PHA depolymerase	*Candidatus* Accumulibacter phosphatis UW-1	CAP2UW1_3195	47	CAP2UW1_3195	47		
TFO_HY_00276340	PhaP	Phasin	*Azospirillum amazonense* Y2	WP_004272223	45				
TFO_HY_00276330	PhaP	Phasin	*Azospirillum amazonense* Y2	WP_004272223	41				
TFO_HY_00271300	PhaR	PHA synthesis repressor	*Thalassospira profundimaris* WP0211	WP_008889414	59				
TFO_HY_00271310		Transcriptional regulator	*Caenispirillum salinarum* AK4	WP_009539279	75				

Table S8. Polyphosphate metabolism genes in TFO71's draft genome with top BLAST hits and top hits in genomes of *Candidatus Accumulibacter phosphatis* UW-1 and *Microlunatus phosphovorus* NM-1.

TFO71 locus	Gene product		Top BLAST hit	Locus	%ID	UW-1 locus	%ID	NM-1 locus	%ID
TFO_HY_00273110	Ppx	Exopolyphosphatase	*Rhodospirillum rubrum* ATCC 11170	Rru_A2160	53	CAP2UW1_1064	34	MLP_44770	38
TFO_HY_00273120	Pap	Polyphosphate:AMP phosphotransferase	*Magnetospirillum magneticum* AMB-1	Amb3969	60	CAP2UW1_4020	50	MLP_05750	38
TFO_HY_00273130	Ppk1	Polyphosphate kinase 1	*Geobacter metallireducens* GS-15	Gmet_0133	62	CAP2UW1_1063	40	MLP_47700	42
TFO_HY_00273140	Php	Signal transduction histidine kinase	*Hyphomicrobium denitrificans* 1NES1	HypdeDRAFT_3037	48	CAP2UW1_2567	33	MLP_39120	38
TFO_HY_00269140	PhoR	Signal transduction histidine kinase	*Oceanibaculum indicum* P24	P24_02926	53	CAP2UW1_1997	38	MLP_47350	35
TFO_HY_00269130	PstS	Phosphate ABC transporter, periplasmic	*Lamprocystis purpurea* DSM 4197	A39ODRAFT_03223	69	CAP2UW1_2005	33	MLP_47720	40
TFO_HY_00269120	PstC	Phosphate ABC transporter, permease	*Rhodospirillum rubrum* S1, ATCC 11170	Rru_A0599	67	CAP2UW1_2004	43	MLP_47730	32
TFO_HY_00269115	PstA	Phosphate ABC transporter, inner membrane	*Caenispirillum salinarum* AK4	C882_3464	74	CAP2UW1_2003	45	MLP_47740	41
TFO_HY_00269110	PstB	Phosphate ABC transporter, ATP-binding	*Nisaea* sp. BAL199	BAL199_11371	74	CAP2UW1_2002	59	MLP_47750	57
TFO_HY_00269100	PhoU	Phosphate ABC transporter regulatory protein	*Rhodocista centenaria* SW	RC1_3266	55	CAP2UW1_1748	46		
TFO_HY_00269090	PhoB	Phosphate regulon transcriptional regulator	*Nisaea* sp. BAL199	BAL199_11361	69	CAP2UW1_1996	51		

Table S9. Relevant *Candidatus Accumulibacter phosphatis* UW-1 gene cassettes and top BLAST hits.

PHA Metabolism		Gene product	Top BLAST hit	Locus	Similarity (%)
CAP2UW1_3190	PhaB	Short-chain dehydrogenase/reductase	*Dechlorosoma suillum* PS	Dsui_1624	75
CAP2UW1_3189	PhaA	Acetyl-CoA acetyltransferase	*Chromobacterium violaceum* ATCC 12472	atoB	64
CAP2UW1_3188	PhaJ	Acyl dehydratase (MaoC domain protein dehydratase)	*Variovorax paradoxus* S110	Vapar_5769	50
CAP2UW1_3187	PhaJ	Acyl dehydratase (MaoC domain protein dehydratase)	Cupriavidus necator HPC(L)	B551_02618	64
CAP2UW1_3186	PhaE	Poly(R)-hydroxyalkanoic acid synthase PhaE subunit	*Candidatus Chlorothrix halophila*	CCHmeta_04057	30
CAP2UW1_3185	PhaC	poly(R)-hydroxyalkanoic acid synthase, class III, PhaC subunit	*Candidatus Chlorothrix halophila*	CCHmeta_04058	49
CAP2UW1_1535	PlsX	phosphate:acyl-[acyl carrier protein] acyltransferase	*Dechlorosoma suillum* PS	Dsui_2445	78
CAP2UW1_1536	FabH	3-oxoacyl-[acyl-carrier-protein] synthase III	*Dechlorosoma suillum* PS	Dsui_2446	76
CAP2UW1_1537	FabJ	[Acyl-carrier-protein] S-malonyltransferase	*Dechlorosoma suillum* PS	Dsui_2447	81
CAP2UW1_1538	FabG	3-oxoacyl-[acyl-carrier-protein] reductase	*Dechlorosoma suillum* PS	Dsui_2448	76
CAP2UW1_1539	AcpP	Acyl carrier protein	*Laribacter hongkongensis* HLHK9	LHK_00569	95
CAP2UW1_1540	FabF	3-oxoacyl-[acyl-carrier-protein] synthase II	*Dechlorosoma suillum* PS	Dsui_2450	76
CAP2UW1_0642	PhaP	Phasin	*Dechloromonas aromatic* RCB	Daro_3423	49
CAP2UW1_0643	PhaP	Phasin	*Dechloromonas aromatic* RCB	Daro_3019	89
CAP2UW1_3191	PhaC	Polyhydroxyalkanoate synthase, Class I	*Polymorphum gilvum* SL003B-26A1	SL003B_0470	57
CAP2UW1_4338	PhaC	Polyhydroxyalkanoate synthase, Class I	*Dechlorosoma suillum* PS	Dsui_0971	56
CAP2UW1_0143	PhaC	Polyhydroxyalkanoate synthase, Class I	*Ruegeria pomeroyi* DSS-3	SPO0112	58
CAP2UW1_0144	FabI	Enoyl-[acyl-carrier-protein] reductase	*Hyphomicrobium denitrificans* 1NES1	Hypde_03280	67
CAP2UW1_0379	FabF	3-oxoacyl-ACP synthase 2	*Dechloromonas aromatica* RCB	Daro_4170	75
CAP2UW1_0368	FabB	3-oxoacyl-ACP synthase 1	*Dechloromonas aromatica* RCB	Daro_4185	73
CAP2UW1_0979	FabH	3-oxoacyl-ACP synthase 3	*Ralstonia solanacearum* Po82	RSPO_m00190	73
CAP2UW1_1898	PhaZ	PHA depolymerase	*Azoarcus* sp. KH32C	AZKH_0291	60
CAP2UW1_3195	PhaZ	PHA depolymerase	*Microvirga* sp. WSM3557	MicloDRAFT_00044920	52
Glycogen metabolism					
CAP2UW1_0254	Glk	Glucokinase	*Neisseria mucosa* C102	HMPREF0604_00119	50
CAP2UW1_0255	GlgP	Glycogen phosphorylase	*Polaromonas naphthalenivorans* CJ2	Pnap_1105	67
CAP2UW1_0256	GlgB	1,4-alpha-glucan branching enzyme	*Rubrivivax gelatinosus* IL144	RGE_29090	62
CAP2UW1_0257	GlgX	Glycogen debranching enzyme	*Rubrivivax benzoatilyticus* JA2	RBXJA2T_03738	57
CAP2UW1_2124	PGI	Glucose-6-phosphate isomerase	*Dechlorosoma suillum* PS	Dsui_1239	63
CAP2UW1_2125	PGM	Phosphoglucomutase	*Gloeobacter violaceus* PCC 7421	gll3983	66
CAP2UW1_2126	GlgC	Glucose-1-phosphate adenylyltransferase	*Polaromonas naphthalenivorans* CJ2	Pnap_1106	71
CAP2UW1_2127	GlgA	Glycogen synthase, ADP-glucose type	*Burkholderia phymatum* STM815	Bphy_1798	55
Polyphosphate					
CAP2UW1_1063	Ppk1	Polyphosphate kinase 1	*Dechlorosoma suillum* PS	Dsui_2814	77
CAP2UW1_1064	Ppx	Exopolyphosphatase	*Dechloromonas aromatica* RCB	Daro_1205	65
CAP2UW1_1995	PhoU	Phosphate uptake regulator	*Nitrosomonas* sp. AL212	NAL212_0182	54
CAP2UW1_1996	PhoB	Phosphate regulon transcriptional regulatory protein	*Rhodoferax ferrireducens* T118	Rfer_0578	57
CAP2UW1_1997	PhoR	Phosphate regulon sensor kinase	*Thiobacillus thioparus* DSM 505	B058DRAFT_02608	48
CAP2UW1_1747	PstS	Phosphate binding protein	*Chloroherpeton thalassium* ATCC 35110	Ctha_0917	52
CAP2UW1_1748	PhoU	Phosphate transport system regulatory protein	*Thermomicrobium roseum* DSM 5159	trd_1827	33
CAP2UW1_1749	PstC	Phosphate ABC transporter, inner membrane subunit	*Chloroherpeton thalassium* ATCC 35110	Ctha_0918	64
CAP2UW1_1750	PstA	Phosphate ABC transporter, inner membrane subunit	*Chloroherpeton thalassium* ATCC 35111	Ctha_0919	59
CAP2UW1_1751	PstB1	ABC transporter-related protein	*Chloroherpeton thalassium* ATCC 35112	Ctha_0920	61
CAP2UW1_1752	PstB2	ABC transporter-related protein	*Chloroherpeton thalassium* ATCC 35113	Ctha_0921	58
CAP2UW1_2010		Mn/Zn ABC transporter, ATPase component	*Microvirga* sp. Lut6	MicLut6DRAFT_00002740	62
CAP2UW1_2009		Zn ABC transporter, inner membrane permease component	*Rhodospirillum rubrum* S1	Rru_A3703	66
CAP2UW1_2008		Mn/Zn ABC transporter, periplasmic component	*Rhodospirillum rubrum* S1	Rru_A3704	56
CAP2UW1_2007	PstS	Phosphate ABC transporter, substrate binding protein SphX	*Nitrosospira multiformis* ATCC 25196	Nmul_A0897	61
CAP2UW1_2006		hypothetical	*Methylophilus* sp. 5	Meth5DRAFT_0447	40
CAP2UW1_2005	PstS	Phosphate binding protein	*Nitrosospira multiformis* ATCC 25196	Nmul_A0897	74
CAP2UW1_2004	PstC	Phosphate ABC transporter, inner membrane subunit	*Dechlorosoma suillum* PS	Dsui_3055	73
CAP2UW1_2003	PstA	Phosphate ABC transporter, inner membrane subunit	*Nitrosomonas* sp. AL212	NAL212_2745	76
CAP2UW1_2002	PstB	ABC transporter-related protein	*Geobacter sulfurreducens* KN400	KN400_1072	78
PiT transporters					
CAP2UW1_3788	PiT	Phosphate transporter (PiT)	*Duganella zoogloeoides* ATCC 25935	F460DRAFT_03046	61
CAP2UW1_2085	PiT	Phosphate transporter (PiT)	*Dechloromonas aromatica* RCB	Daro_0679	79
CAP2UW1_3785	PiT	Phosphate transporter (PiT)	*Thiorhodovibrio* sp. 970	Thi970DRAFT_1234	63
CAP2UW1_3733	PiT	Phosphate transporter (PiT)	*Arthrospira platensis* C1	SPLC1_S201490	52

Table S10. Relevant *Microlunatus phosphovorus* NM-1 gene cassettes and top BLAST hits.

Glycogen					
MLP_43000	GlgC	Glucose-1-phosphate adenylyltransferase	*Propionibacterium acidipropionici* DSM 4900	YWODRAFT_00225	72
MLP_43020	GlgA	Glycogen synthase	*Propionibacterium acidipropionici* DSM 4900	YWODRAFT_00224	72
MLP_46800	GlgP	Glycogen phosphorylase	*Cellulomonas flavigena* 134	Cfla_0539	77
MLP_36670	GlgX	Glycogen debranching enzyme	*Cellulomonas flavigena* 134	Cfla_1743	71
MLP_29690	GlgX	Glycogen debranching enzyme	*Propionibacterium acnes* TypeIA2 P.acn17	TIA2EST22_05540	64
MLP_29700	GlgE	4-alpha-glucanotransferase	*Dehalobacter* sp. FTH1	A37GDRAFT_01291	63
MLP_29710	TreS	Alpha amylase	*Propionibacterium freudenreichii shermanii* CIRM-BIA1	PFREUD_10640	46
MLP_29720	GlgB	1,4-alpha-glucan branching protein	*Dehalobacter* sp. FTH1	A37GDRAFT_01289	70
PolyP					
MLP_47680	GlnR	Glutamine synthesis transcriptional regulator	*Kribbella flavida* DSM 17836	Kfla_0973	65
MLP_47690	MshD	Mycothiol acetyltransferase	*Isoptericola variabilis* J5	IsoJ5_00004080	43
MLP_47700	Ppk	Polyphosphate kinase	*Dehalobacter* sp. FTH1	A37GDRAFT_03120	77
MLP_47710	MutT	NTP pyrophosphohydrolase	*Dehalobacter* sp. FTH1	A37GDRAFT_03119	40
MLP_47720	PstS	Phosphate ABC transporter phosphate-binding protein	*Clavibacter michiganensis michiganensis* NCPPB 382	CMM_2505	56
MLP_47730	PstC	Phosphate ABC transporter permease	*Clavibacter michiganensis sepedonicus* ATCC 33113	CMS_2726	73
MLP_47740	PstA	Phosphate ABC transporter permease	*Candidatus* Aquiluna sp. IMCC13023	IMCC13023_01210	66
MLP_47750	PstB	Phosphate ABC transporter ATP-binding protein	*Arthrobacter phenanthrenivorans* Sphe3	Asphe3_03510	79
MLP_50300	Ppk2	Polyphosphate kinase 2	*Cellulomonas fimi* NRS 133	Cfi_0001.00012520	82
MLP_05750	Ppk2	Polyphosphate kinase 2	*Nakamurella multipartita* Y-104	Namu_3897	84
MLP_26610	PpgK	Polyphosphate-dependent glucokinase	*Nakamurella multipartita* Y-104	Namu_3763	62
MLP_05430	PpgK	Polyphosphate-dependent glucokinase	*Dehalobacter* sp. FTH1	A37GDRAFT_02973	71
MLP_44770	Ppx	Exopolyphosphatase	*Propionibacterium avidum* ATCC 25577	HMPREF9153_1563	51
PHA					
MLP_12780	PhaJ	enoyl-CoA hydrolase	*Burkholderia* sp. JPY251	B020DRAFT_01576	82
MLP_23080	YfcY	beta-ketothiolase	*Kribbella flavida* DSM 17836	Kfla_3736	66
MLP_23090	YfcX	Enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, 3-hydroxyacl-CoA epimerase	*Actinopolymorpha alba* DSM 45243	B141DRAFT_03809	79
MLP_02560	MPPKEI	3-oxoacyl-ACP synthase 2	*Haloglycomyces albus* DSM 45210	HalalDRAFT_1203	64
MLP_33250	MSPTPI	3-oxoacyl-ACP synthase 2	*Dehalobacter* sp. FTH1	A37GDRAFT_01879	70
MLP_42120	MTGTR	3-oxoacyl-ACP synthase 3	*Arthrobacter arilaitensis* re117	AARI_28540	68
MLP_50950	MSGNA	3-oxoacyl-ACP synthase 3	*Kytococcus sedentarius* 541	Ksed_07160	58