

A systems biology framework to identify microbial strain-level interactions in engineered ecosystems

By
Francisco Moya-Flores

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
(Civil and Environmental Engineering)

at the
UNIVERSITY OF WISCONSIN- MADISON
2019

Date of final oral examination: January 16th, 2019

The dissertation is approved by the following members of the Final Oral Committee:

Katherine D. McMahon, Professor, Civil and Environmental Engineering

Daniel R. Noguera, Professor, Civil and Environmental Engineering

Gregory Harrington, Professor, Civil and Environmental Engineering

Timothy J. Donohue, Professor, Bacteriology

Acknowledgments

I would like to express my gratitude to all the people that helped me to conceive this research and provided me with thoughtful discussions during the course of my research here in UW-Madison. Especial thanks to my mentor, Dr. Katherine McMahon, who gave me the opportunity to join UW-Madison, inspired me and walked me through developing a rational story for my research, and taught me a path that has become my career as an Environmental Engineer. I cannot be more grateful for her kindness and understanding.

Thanks to Dr. Daniel Noguera, Dr. Greg Harrington and Dr. Timothy Donohue for serving on my Ph.D. committee. Their willingness to help to develop my curriculum in UW-Madison encouraged me to pursue this investigation. Their excellence in research and teaching has profoundly impacted my formation in academia. I would like to acknowledge Dr. Noguera, who received me as a participant on his laboratory meetings, generously invested his time and genuinely provided me with deep insights into my research.

I would also like to thank my fellow lab-mates of the McMahon Lab and Noguera Lab group for their kindness, advice, and support. This includes Sarah Stevens, Robin Rohwer, Travis Korosh, Andrew Dutcher, Joshua Hamilton, Shaomei He, Cristina Herren, Christopher Lawson, Ben Peterson, Alexandra Linz, Pamela Camejo, Natalie Keene, Ben Oyserman, Matthew Scarborough, among others. Especial thanks to Ben Oyserman. Without his support and collaboration, this research would not have been possible. He not only helped me get started, but was

consistently available to discuss my research, provide his advice, and share his expertise in order to help this project be successful. Many thanks to Jacqueline (Jackie) Bastyr Cooper for her assistance in the lab with training, protocols, and equipment. The maintenance of the reactors, sampling, data collection and interpretation would not have been possible without the help of more than a dozen undergraduate students. This includes Aldo Ventura, Douglas Chalmers, Daniel Vigil, Tenzin Khangkar, Char'Lee King, Coty Weathersby, Claudia Santana, Kyle Winkler, Brandon Nutley, Frances Palmer, Shelby Sample, Diana Mendez, Michael Eng Khor, and Matthew Kizaric.

Special thanks to Becas Chile Scholarship for the funding provided to my graduate studies, as well as the National Science Foundation, for their accessibility to continuous funding in research, and the Joint Genome Institute and UW-Madison Biotechnology Center for DNA sequencing services provided for this study.

I would also like to express my greatest regards to my family. My parents, for their hope and unconditional support. They gave me even what was beyond their means to help me become a professional. To my brother, from whom I learned that effort returns priceless rewards. And to my wife, Pamela Camejo, I dedicate this work. Her support and perseverance were key in succeeding in my journey in graduate school.

Table of Contents

Acknowledgments	i
Table of Contents	iii
List of Figures	viii
List of Tables	x
Chapter 1: Introduction	1
Overview of wastewater treatment	1
Enhanced Biological Phosphorus Removal.....	2
Accumulibacter	3
Genome-scale metabolic reconstruction and constraint-based modeling.....	5
Bacteriophages and their role in ecosystem fitness	6
Accumulibacter defense mechanisms against phage predation	7
Thesis outline	9
Chapter 2: Uncovering the driving force of bacteriophages within the diversity of Accumulibacter phosphatis through CRISPR-Cas	12
Abstract	13
Main	14
Methods	17
2.1 Reactor operation and sample preparation	17
2.2 Accumulibacter clades quantification.....	17
2.3 Library preparation and DNA sequencing.....	18
2.4 Raw read processing and de novo assembly of metagenomic reads	18

2.5 CRISPR amplicon sequencing	20
2.5.1 Construction and Sequencing of Custom Amplicon Libraries	20
2.6 Locus architecture characterization	22
2.7 CRISPR immune structure	22
2.8 EPV1 phage quantification	23
Results	25
Metagenomic-based evidence for differential acquisition of CRISPR spacers at the leader end.....	26
CRISPR loci dynamics reveal intraspecies diversity.....	26
Discussion.....	29
Acknowledgements.....	33
Figure descriptions	34
Figures	36
Supplementary Text.....	39
Supplementary Materials	40
Supplementary Figures.....	40
Supplementary Tables	41
Chapter 3: Metabolic plasticity of two co-occurring Accumulibacter strains revealed through metatranscriptomics	54
Abstract	55
Main	56
Methods	60
3.1 Bioreactor operation	60
3.2 Experimental design	60
3.3 DNA and RNA extractions.....	61
3.4 Accumulibacter clades quantification.....	61

3.5 Library preparation and DNA sequencing	62
3.6 Construction of Prokaryotic Illumina RNA Libraries	62
3.7 Raw read processing and de novo assembly of metagenomic reads	63
3.8 Genome annotation and metabolic reconstruction	64
3.9 Identification of orthologous genes	64
3.10 Definition of functional gene sets	65
3.11 Metatranscriptomics analysis	65
3.12 Differential gene expression calculations	65
Results	67
Metagenomic sequencing and binning of Accumulibacter genomes	67
Metatranscriptomics profiling of Accumulibacter clades under classical EBPR conditions	68
Acetate conversion/transport and PHB synthesis.....	70
Anaerobic TCA operation.....	72
Respiration.....	73
Denitrification capabilities	74
Discussion.....	76
Clade-specific responses to anaerobic carbon uptake.....	77
Importance of substrate concentration during polymer storage	79
Concluding remarks.....	79
Acknowledgements.....	81
Figures Descriptions	82
Figures	84
Tables	89
Supplementary Materials	90
Supplementary Figures.....	90
Supplementary Tables	92

Chapter 4: iCAP366: A genome-scale metabolic network reconstruction of the archetypal PAO *Accumulibacter phosphatis*.....101

Abstract	102
Main	103
Accumulibacter	104
Lack of consensus on Phenotypic Potentials.....	105
Current State of EBPR Modeling	106
Motivation	106
Methods	108
4.1 Computational refinement.....	108
4.2 COBRApy Model Creation.....	108
4.3 Defining Anaerobic Constraints and Objectives	109
4.4 Simulation of the Acetate and Glycogen flux ratio.....	109
4.5 Comparison of Measured and Simulated Data.....	110
4.6 Estimated PHA and Pi Stoichiometry	110
Results	111
Creation of a standard anaerobic model.....	112
Simulation of Anaerobic Carbon Utilization	112
Resolution of FBA results and its validation.....	114
Estimation of stoichiometric constants.....	114
Discussion.....	115
Lack of Evidence for an Active Calvin Cycle During Acetate Uptake	115
Hydrogen Production Expands PHA Production Space	116
A Complete Anaerobic TCA Cycle Could Be a Feasible Phenotype	117
Future Directions.....	118
Acknowledgements.....	120
Figures Descriptions	121

Figures	122
Tables	126
Supplementary Material	129
Chapter 5: Recommendations for Future Research.....	131
Use of metatranscriptomics constraints in genome-scale modeling	131
Transcriptional Regulatory Networks	134
References cited	135

List of Figures

Chapter 1

Figure 1 A summary of the significant features of the biochemical models for PAO.	3
Figure 2 Maximum likelihood tree of <i>ppk1</i> genes from the Accumulibacter lineage.	4

Chapter 2

Figure 1 Metagenomic-based evidence for differential acquisition of CRISPR spacers at the leader end.	36
Figure 2 CRISPR loci dynamics reveal an intraspecies diversity.	37
Figure 3 Fitness consequences of generating spacer diversity.	38

Chapter 3

Figure 1 Accumulibacter Pan-transcriptome in response to classic feast and famine conditions.	84
Figure 2 Functional differences of gene expression at the clade level.	85
Figure 3 Gene expression dynamics across the EBPR cycle.	86
Figure 4 Presence and expression of denitrification genes across Accumulibacter Clades IA-UW3 and IIC-UW6.	87
Figure 5 Proposed mode of operation for anaerobic acetate uptake and conversion to PHB in Accumulibacter.	88
Supplementary Figure S1 Time-series EBPR data.	90

Supplementary Figure S2 ANI vs Percent Alingment of several Accumulibacter genomes.....	91
---	----

Supplementary Data Table S4 (next 6 pages) Gene expression dynamics analyzed under this study.	93
---	----

Chapter 4

Figure 1 Simulation of Accumulibacter's phenotypic response to varying glycogen to acetate flux ratios under "standard" constraints.	122
---	-----

Figure 2 Flux profiles of proposed anaerobic mechanisms.	123
---	-----

Figure 3 Resolution of FBA simulations with previously obtained results.....	124
--	-----

Figure 4 Model estimated PHA production and Pi release per 1 Cmmol of acetate.	125
---	-----

List of Tables

Chapter 2

Supplementary Table S1. Metagenomic sequencing statistics	41
Supplementary Table S2. Reference sequences used in this study	41

Chapter 3

Table 1 Metagenome-assembled genomes (MAGs) of Accumulibacter obtained from lab-scale reactors in 2016.....	89
Supplementary Table S1 Metatranscriptomics reads analysis.....	92
Supplementary Table S2 Accumulibacter Pan-transcriptome information.....	92
Supplementary Table S3 mRNA reads mapped to all four Accumulibacter clades present in the chemostat.....	93

Chapter 4

Table 1 Summary of iCAP366 GEM.	126
Table 2 Pathways constraints used during this study.	127
Table 3 All possible scenarios defined by combining the selected phenotypes and selected constraints.	128
Supplementary Table S1 Maximum Polymer and Pi fluxes obtained for each simulation.	129

Chapter 1: Introduction

Overview of wastewater treatment

Culture-independent methods available for probing mixed microbial communities have drastically increased our ability to gather comprehensive molecular-level information on an ever-growing number of microorganisms. These communities contribute to diverse aspects of the environment, varying from nutrient cycling to engineered ecosystem processes. Bioreactors designed to simulate the activated sludge process are ideal for investigations on the system biology principles in natural ecosystems due to the high microbial density and the discrete operational parameters that may be monitored over long periods ¹. These systems employ uncultivated microbial communities to efficiently remediate wastewater streams before being discharged to the aquatic environment. A central idea in the activated sludge process is that microbially-mediated processes for the removal of C, N, and P are hierarchically organized and that this structure plays an essential role in the system's dynamics. These communities display unique characteristics that have been primarily exploited during the past decades and their specialized role during wastewater treatment can be classified by substrate specificity. However, despite the widespread usage of this biotechnology, process performance can be transiently or chronically poor, presumably because of unstable microbial community structure or activity ².

Enhanced Biological Phosphorus Removal

Enhanced Biological Phosphorus Removal (EBPR) is a variant of the activated sludge wastewater treatment process; the most widely used environmental biotechnology system worldwide. EBPR is a biochemically complex process achieved through enrichment of a group of bacteria known as Polyphosphate Accumulating Organisms (PAO) in activated sludge employing alternating anaerobic and aerobic conditions. Under anaerobic conditions, PAO take up volatile fatty acids (VFAs) and convert them to polyhydroxyalkanoates (PHA), using energy obtained from the hydrolysis of two internally stored polymers, polyphosphate (poly-P) and glycogen. Poly-P is hydrolyzed to orthophosphate and released from PAO cells, while glycogen is converted to PHA and CO₂. Aerobically, PAO oxidize PHA to gain energy for growth, glycogen replenishment and phosphorus uptake³⁻⁵. Phosphorus removal is achieved through the wastage of excess sludge with high poly-P content.

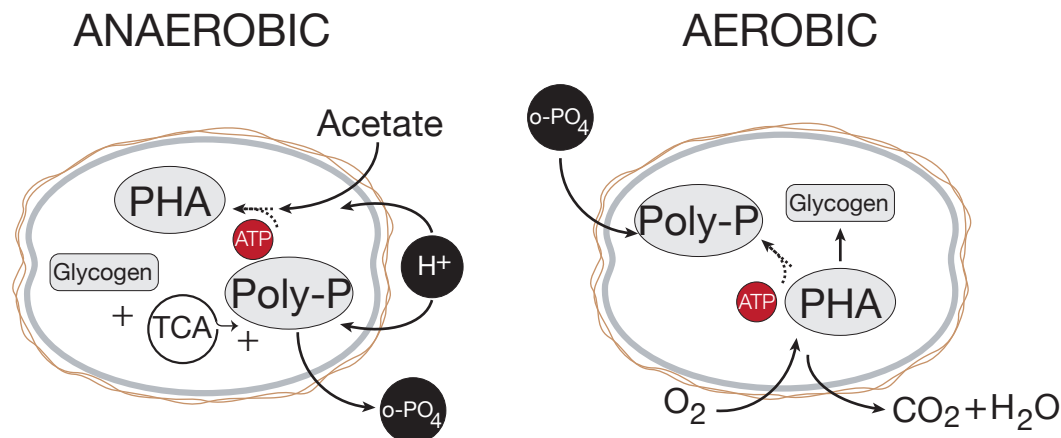


Figure 1 | A summary of the significant features of the biochemical models for PAO.

In the anaerobic phase, acetate is stored as PHA. PHA production requires energy (ATP) and reducing power (NAD(P)H). ATP (in red) is supplied by polyphosphate degradation and glycogen degradation. In the aerobic phase, when oxygen is available for respiration, the PHA reserves of PAO ensures their dominance in the SBR microbial ecosystem. The restoration of polyphosphate reserves via ATP depletes the water of P_i, thus giving rise to EBPR.

Accumulibacter

In wastewater treatment plants as well as in lab-scale acetate-fed EBPR reactors, the dominant organism is a member of the Betaproteobacteria in the Rhodocyclus group, named *Candidatus Accumulibacter phosphatis* (hereafter referred to as Accumulibacter) ^{6,7}. No pure culture of this organism is yet available, though culture-independent molecular techniques are providing much of the essential information that historically could only be obtained using pure cultures ^{8,9}. These studies have revealed that Accumulibacter is subdivided into two main Types (I and II), each of which contains several coherent clades (Figure 2) ^{8,10,11}. With the identification of these clades within the Accumulibacter lineage, questions have arisen as to whether these clades play distinct roles in EBPR.

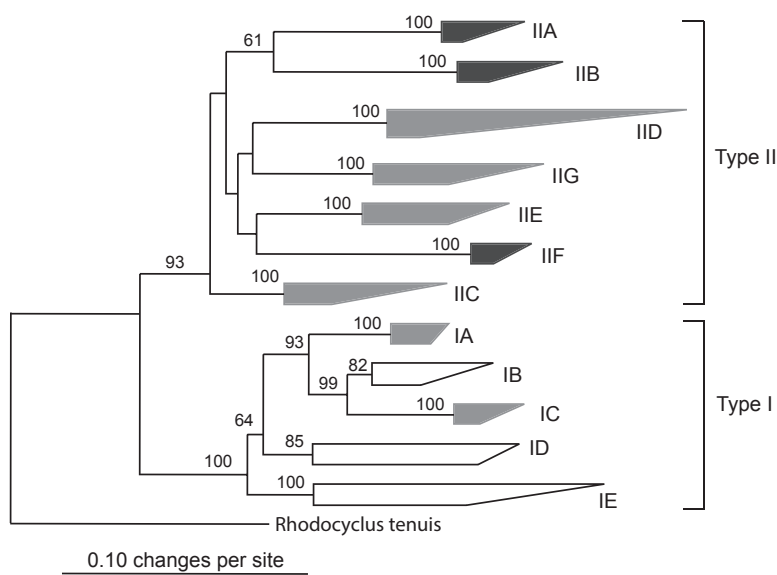


Figure 2 | Maximum likelihood tree of *ppk1* genes from the *Accumulibacter* lineage.

Accumulibacter is comprised of two Types (I and II indicated by brackets), each consisting of many of monophyletic clades. Clades in black are exclusively derived from EBPR systems and environmental samples collected near EBPR treatment plants. Clades in grey contain sequences from both EBPR systems and natural habitats not associated with EBPR treatment plants. Clades in white exclusively consist of environmental samples, particularly from an estuary sediment ^{8,10,11}.

Several researchers have investigated kinetics and substrate specificity of *Accumulibacter*, but these studies did not discriminate between clades, which appear to have ecologically distinct characteristics. Preliminary data obtained from our laboratory indicate that clade IA has higher acetate uptake rates and higher phosphate release rates, and that clade IA can reduce nitrate while clade IIA cannot ¹². These results suggest that *Accumulibacter* clades inhabit different niches in EBPR ecosystems, each providing an essential role in ecosystem function, and their differences could offer a competitive advantage to one clade over another, in a

mixed community, depending on the operating conditions. Moreover, the genomic architecture that underlies ecological divergence and speciation with gene flow is still unidentified for *Accumulibacter*, encouraging us to study significant differences in gene content for both clades.

Genome-scale metabolic reconstruction and constraint-based modeling.

Two decades ago, a constraint-based model (CBM) was used to study EBPR in a laboratory scale Sequencing batch reactor (SBR). The model could make predictions that were in reasonable numerical agreement with experimental observations. A much more accurate metabolic model can now be constructed using the *Accumulibacter* clade genomes and expression profile easily accessible due to the advent of metagenomics, metatranscriptomics, and metaproteomics approaches, with each iteration building on previous work while incorporating new knowledge. The ability to sequence whole genomes has made possible to formulate CBMs at the genome-scale and allowed representation of the complete metabolic gene content in the assessment of phenotypic functions. As the generation of omics data became cheaper and as larger data sets appeared, researchers began to incorporate these data sets into CBMs. Flux balance analysis (FBA) is a form of CBM adapted by the systems biology community and it has been established as a leading approach for studying cellular metabolism, enabling the explicit and quantitative description of metabolic network and imposes mass balance as the canonic constraint.

Bacteriophages and their role in ecosystem fitness

Natural phage communities are reservoirs of the greatest uncharacterized genetic diversity on Earth ¹³, and therefore, phage-host interactions represent one of the most dominant biological interactions. Phages primarily act as a predator, entering host cells to replicate and lysing them to propagate new phage particles. The critical role of phage in structuring microbial communities was first recognized in the oceans, where they appear to influence biogeochemical cycles globally, provide and regulate microbial biodiversity, cycle carbon through marine food webs, and to control bacterial population explosions ¹⁴.

Comparative genomics and bioinformatics serve as useful tools for identifying phage-bacteria in the activated sludge process. For example, previous studies showed the presence of viral elements in EBPR reactors, likely belonging to phages infecting at least one member of the *Accumulibacter* lineage ¹⁵. This phage-host interaction is thought to recreate the 'kill the winner' hypothesis, in which hosts that become abundant due to uptake efficiency become targets of viral attack ^{16,17}. This interaction results in boom-bust cycles of phage and host abundance, in which uninfected host populations grow until they are infected and lysed, with associated exponential growth and collapse of phage abundance.

Identification of a lysogenic bacteriophage capable of infecting and transducing genetic information into the *Accumulibacter* lineage could represent a significant contribution to the field. However, to date, no phage capable of infecting members of this group has been confirmed. The factors responsible for determining which

species-like clade dominates the sludge are unknown but could include phage-mediated population crashes. Here, phages can impart frequency-dependent selection on bacteria that succeed in having higher population densities, which are also those most susceptible to phage attack, both due to the higher frequencies and because of trade-offs those bacteria may have made regarding phage susceptibility to grow to these higher densities ¹⁸.

Accumulibacter defense mechanisms against phage predation

A primary tool employed by microbes to avoid predation is to prevent phage infection by the production of extracellular polymeric substances (EPS). EPS can provide a first line of defense against phage predation by masking attachment sites on the cell surface¹⁹. The observed redundancy and variability of EPS gene cassettes in both *Accumulibacter* genomes may impede strain-specific targeting of EPS by phage ²⁰. A recently discovered adaptive immune mechanism in bacteria and archaea are Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated with Cas genes. The CRISPR-Cas system functions as a prokaryotic immune system conferring resistance to foreign genetic elements such as plasmids and phages, by storing segments of previously confronted phage or other external DNA as spacers that are surrounded by repeat sequences adjacent to a series of Cas genes ^{21,22}. Previous research has also revealed the presence of CRISPR-Cas systems in *Accumulibacter* ^{15,20}. Further, the analysis of these elements on both clades suggested their immunity to phage was unequal, due to differences on their CRISPR spacer elements. These elements could affect genome stability and

population dynamics and as such are likely to be relevant for species survival, biological activity, and diversity, and eventually also for the performance of phosphorus removal.

Thesis outline

Little is known about the biological factors that affect genome stability in EBPR populations or how these perturbations may, in turn, affect biological fitness and species dynamics during phosphorus removal. In this thesis, I describe the *Accumulibacter* lineage from a 'systems biology' approach. This work is divided into three principal manuscripts, and they are aimed to tackle the most controversial features of *Accumulibacter* at the genome level. These manuscripts are described below.

Chapter 2: The purpose of this investigation was aimed to solve the following questions: Are there patterns in how interactions differ in the selection pressures that they take place on bacterial species and their infecting phages? Are there patterns in how these interactions change over evolutionary time? And, are some interactions and ecological conditions more likely to lead to coevolution than others? Host-parasite coevolution plays an essential role in shaping genotypic, phenotypic and community-level diversity ²³. In this study, I evidenced host adaptation to phage as well as its adaptation to host in enriched communities performing phosphorus removal. Here I developed a novel DNA sequencing application to determine the micro-diversity of *Accumulibacter* spp. IA-UW3 by studying CRISPR spacer dynamics from a bioreactor continuously operating for over a decade. Relatively few experimental coevolution studies have attempted to increase community complexity beyond pairwise interactions.

Chapter 3: One of the discoveries brought by the comparative genomics study performed on *Accumulibacter* is the prediction of significant lateral gene transfer occurred in the past ^{15,24,25}. In this chapter, I assembled several genomes from *Accumulibacter* to reveal the dynamic patterns between members of the community. A comprehensive metatranscriptomics profiling was analyzed to discover differences at the gene expression levels across the canonical feast-famine conditions of the EBPR cycle between two *Accumulibacter* strains.

Chapter 4: In this work, we ran FBA to simulate the metabolic capabilities of *Accumulibacter phosphatis* under the regular EBPR cycle. We employed omics data integration as constraints to create, validate, calibrate and improve the predictive powers of metabolic models of *Accumulibacter*.

Intentionally blank page

Chapter 2: Uncovering the driving force of bacteriophages within the diversity of *Accumulibacter phosphatis* through CRISPR-Cas

Francisco Moya-Flores, Ted Kim, Daniel R. Noguera, Rachel Whitaker and
Katherine D. McMahon.

Experiment design: FMF and KDM

DNA processing and library preparation: FMF

Data analysis: FMF, TK and KDM

Manuscript writing: FMF, TK, DRN, RW and KDM

Abstract

Bacteria have evolved sophisticated phage defenses, given the abundance of viruses and their ability to attack abundant bacterial populations. One such mechanism involves adaptive immunity in bacteria and archaea mediated by the CRISPR-Cas system. Although the molecular mechanisms of this process have been extensively studied in pure cultures, little is known about the extent or role of coevolution in mixed microbial populations, in which strain diversity limits the ability to develop robust coevolutionary models in the context of the CRISPR-Cas system. Here we show that a self-assembled community can form a complex structure where the host can overcome viral predation as a result of immunity conferred by diversified CRISPR repeat-spacer alleles. As a population these diversified alleles prevent virus invasion and stabilize the host diversity. We suggest that CRISPR diversity may be maintained within the bacterial population due to distributed immunity among individuals that have independently acquired spacers in their CRISPR loci. Finally, we discuss the evolutionary impacts of distributed immunity in a natural population.

Main

Bacterial viruses, or phages, are highly abundant among microbial communities and act in a density-dependent manner. The underlying processes governing prokaryotic interference against the invasive nucleic acids introduced by phages and the corresponding evolution of phage defenses are unresolved questions in microbial ecology. One defense mechanism in bacteria and archaea is adaptive immunity, which is associated with Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) linked to CRISPR associated (Cas) genes. The CRISPR-Cas system confers resistance to foreign genetic elements by storing segments of DNA from previously confronted phages as spacers between arrays of repeat sequences ^{21,22}. The acquisition and incorporation of spacers of viral DNA into the host protect them from subsequent infection. Conversely, viruses evade host CRISPR-mediated immunity through mutation in homologous protospacers (phage DNA sequences that match CRISPR-spacers) or protospacer-adjacent motifs (PAMs) ^{26,27} and by carrying sophisticated anti-CRISPR mechanisms to overcome CRISPR-Cas ²⁸.

The dynamics and consequences of coevolution between phage and their hosts have been extensively studied with pure cultures ^{29–32}. Similarly, there is a plethora of studies that model CRISPR-Cas diversity via computational simulations ^{33–36}. However, little is known about the extent or role of host-phage coevolution in mixed microbial populations. Recent metagenomic studies of bacteria, archaea, and their viruses provide compelling support that CRISPR-Cas systems play an active role in

the population dynamics and coevolution of these organisms in natural communities^{26,32,37,38}. Bioreactors designed to simulate wastewater treatment are ideal for investigating coevolution due to the high microbial density of their enriched communities and the discrete operational parameters that may be monitored over long periods of time.³⁹

Enhanced Biological Phosphorus Removal (EBPR), a variant of the activated sludge wastewater treatment process, has been used as a model for studying phage-bacteria interactions^{20,39–41}. EBPR is a biochemically complex process achieved through enrichment of a bacterial group known as Polyphosphate Accumulating Organisms (PAO) employing alternating anaerobic and aerobic conditions^{4,6}. The dominant organism in many of these enrichments is a member of the Betaproteobacteria in the Rhodocyclaceae family, named *Candidatus Accumulibacter phosphatis*^{6,7}. No pure culture of this organism is yet available, though culture-independent molecular techniques provide much of the essential information that traditionally has been obtained using pure cultures^{8,9,15,25,42–45}.

In a previous comparative genomic analysis from the same bioreactor used in this study, Flowers et al.¹⁵ revealed the extensive presence of elements derived from phages, plasmids and mosaic mobile genetic elements. Specifically, a nearly complete phage genome was found in a metagenomic assembly⁵. The same phage genotype was assembled from a supernatant-derived metagenome and named EPV1⁴⁰. Thus, the phage is inferred to display a lytic lifestyle, infecting at least one

member of the *Accumulibacter* lineage. Flowers et al.¹⁵ also found diverging CRISPR profiles in two *Accumulibacter* genomes assembled from metagenomic samples. Furthermore, by comparing the CRISPR spacer sequences across both genomes, they concluded that their immunities to phages and plasmids were unequal, despite the fact that these two *Accumulibacter* strains are closely related and perform the same function in the bioreactor. Adaptation of *Accumulibacter* spp. across geographically separated bioreactors was earlier investigated by metagenomic sequencing²⁰. Despite evidence for high dispersal among the sites and relatively little divergence between bacterial genotypes, the CRISPR sequences appeared to be rapidly diverging. These results, which support a model of local phage-mediated selection, suggest that natural bacterial host populations can respond to phage selection through evolved adaptive immunity.

With the evidence obtained from previous metagenomics efforts and the interest to evaluate *Accumulibacter*'s diversity in response to interactions with phages, we analyzed *Accumulibacter*'s CRISPR immune structure by sequencing DNA samples collected from our bioreactor community longitudinally during nine years of continuous operation. Using a combination of whole genome sequencing, CRISPR-targeted amplicon sequencing, quantitative PCR, and fluorescence *in situ* hybridization (FISH), we demonstrate temporal virus-host dynamics in this engineered community.

Methods

2.1 Reactor operation and sample preparation

A laboratory-scale sequencing batch reactor (SBR) was seeded with activated sludge from the Nine Springs Wastewater Treatment Plant located in Madison, Wisconsin. The SBR was operated in four daily cycles of 6 hours with a hydraulic residence time (HRT) and solids residence time (SRT) of 12 hours and four days, respectively, as previously described^{1,5,12}. The reactor was operated for 3636 days, or 909 generations (calculated from the bioreactor's SRT), from inoculation through to the end of the current study. Biomass samples for DNA analysis were collected by centrifuging 2 ml of mixed liquor at 8,000 x g for 3 min, and the resulting pellets were transferred to a -80°C freezer immediately for long-term storage. DNA was extracted from previously frozen biomass using the PowerSoil DNA Extraction Kit (MoBio, Carlsbad, CA) following the manufacturer's instructions.

2.2 Accumulibacter clades quantification

To determine the relative proportions of two *Accumulibacter* clades that were simultaneously enriched in the bioreactor (Clade IA-UW3 and IIA-UW1¹), we used quantitative Polymerase Chain Reaction (qPCR) using clade-specific primers targeting the polyphosphate kinase (*ppk1*) gene and protocols as described by Camejo et al.⁴⁵ We selected nine DNA samples for metagenomic sequencing based on their enrichment in either *Accumulibacter* clade IA or IIA. A phenol-chloroform bead-beating extraction method⁵ was performed to obtain total genomic DNA from

the biomass in preparation for either shotgun metagenomic sequencing or CRISPR-amplicon sequencing.

2.3 Library preparation and DNA sequencing

Metagenomic DNA sequencing on the nine chosen samples was performed at the Department of Energy Joint Genome Institute (Walnut Creek, CA, USA). Briefly, 100 ng of DNA was sheared to 300 bp using the Covaris LE220 (Covaris) and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A- tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA Biosystems). The libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche Light Cycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq2500 sequencer using TruSeq SBS sequencing kits, following a 2x150 indexed run recipe. Raw data for this study consisted of ~100 million 150-bp Illumina HiSeq reads with about ~15 Gpb per sample (Table 1 – supplementary material).

2.4 Raw read processing and de novo assembly of metagenomic reads

For the nine metagenomes obtained from the sequencing effort described above, FASTQ files were quality filtered and read-trimmed using the Sickle software v1.33

(<https://github.com/najoshi/sickle>). Reads were merged with FLASH v1.0.3 ⁴⁶ with a mismatch value of ≤ 0.25 and a minimum of ten overlapping bases from paired sequences, resulting in merged read lengths of 150 to 290 bp. FASTQ files were then converted to FASTA format using the Seqtk software v1.0 (<https://github.com/lh3/seqtk>). Quality-filtered, trimmed and merged metagenomic reads were assembled using the Velvet assembler ⁴⁷ with a k-mer size of 65, a minimum contig length of 200 bp, and a paired-end insert size of 300 bp. Metavelvet ⁴⁸, an extension of Velvet assembler to de novo metagenome assembly from short sequence reads, was used to improve the assembly generated by Velvet.

Genome binning was performed using Maxbin, an automated binning method to recover individual draft genomes from metagenomes using an expectation-maximization algorithm after the assembly of metagenomic sequencing reads⁴⁹. Genome completeness for the resulting bins was estimated using CheckM, an automated method for assessing the quality of a genome using a broader set of marker genes specific to the position of a genome within a reference genome tree and information about the co-location of these genes ⁵⁰. From day 3537 of operation, a metagenome-assembled genome (MAG) was retrieved and taxonomically assigned to *Accumulibacter* clade IA via both ANI comparisons and *ppk1* phylogenetic assignments. Contigs were scaffolded using Medusa ⁵¹, manually inspected for short contigs removal and decontaminated using the JGI tool Prodege ⁵² and anvi'o ⁵³.

2.5 CRISPR amplicon sequencing

All obtained metagenomic reads were mapped against the CRISPR-Cas locus from the *Accumulibacter* clade IA reference genome using *bbmap* (<https://sourceforge.net/projects/bbmap/>). Coverage was visualized using a custom R script available in GitHub (<https://github.com/McMahonLab/>). We used this approach to identify the hypervariable region within all loci. This region was represented by a decrease in coverage or inconsistency on reads mapping to the reference CRISPR locus. We then designed primers to capture this region in depth by following a custom amplicon-sequencing strategy. The primer set CRISPR_CAPIA-leader (5'-CGTTTGTCATGCCGTTTCGTT), and CRISPR_CAPIA-leader R (5'-CGGGGAACGCAGACTCAAG) and Illumina adapter sequences were synthesized. Metagenomic libraries were generated by multiplexing 106 DNA samples retrieved once a month during a 10-yr timespan and sequenced on the Illumina MiSeq platform following a 2x300 indexed run recipe.

2.5.1 Construction and Sequencing of Custom Amplicon Libraries

Purified CRISPR DNA amplicons were submitted to the University of Wisconsin-Madison Biotechnology Center. DNA concentration was verified fluorometrically using either the Qubit® dsDNA HS Assay Kit or Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). Samples were prepared in a similar process to the one described in Illumina's 16S Metagenomic Sequencing Library Preparation Protocol, Part #15044223 Rev. B (Illumina Inc., San Diego, California, USA) with the following modifications: We used the metagenomic-

assembled CRISPR locus obtained during 5/13/2013 to identify the hypervariable region within all loci. This region was represented by a decrease in coverage or inconsistency on reads mapping to the reference CRISPR locus. The CAPIA CRISPR hypervariable region was amplified with fusion primers (forward primer: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTTTGCATGCCGTTTCGTT-3', reverse primer: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGGGGAACGCAGACTCAAG-3'). Region-specific primers were modified to add Illumina adapter overhang nucleotide sequences to the region-specific sequences. Following initial amplification, reactions were cleaned using a 0.7x volume of AxyPrep Mag PCR clean-up beads (Axygen Biosciences, Union City, CA). Using the initial amplification products as template, a second PCR was performed with primers that contain Illumina dual indexes and Sequencing adapters (Forward primer: 5'-AATGATACGGCGACCAACGAGATCTACAC[55555555]ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3', Reverse Primer: 5'-CAAGCAGAAGACGGCATACGAGAT[77777777]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3', where bracketed sequences are equivalent to the Illumina Dual Index adapters D501-D508 and D701-D712, N716, N718-N724, N726-N729). Following PCR, reactions were cleaned using a 0.7x volume of AxyPrep Mag PCR clean-up beads (Axygen Biosciences). Quality and quantity of the finished libraries were assessed using an Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA) and Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific),

respectively. Libraries were pooled in an equimolar fashion and appropriately diluted before sequencing. Paired-end, 300 bp sequencing was performed using the Illumina MiSeq Sequencer and a MiSeq 600 bp (v3) sequencing cartridge. Images were analyzed using the standard Illumina Pipeline, version 1.8.2.

2.6 Locus architecture characterization

An in-house python script was developed to extract spacers from amplified CRISPR loci. Spacers were mapped along loci using repeat locations, identified using BLASTN with a e-value cutoff of 0.01 and window size of 7. Spacers were extracted with the blast+ package `blastdbcmd`⁵⁴. A custom script was used to bin spacers 3 SNPs apart to correct for sequencing error or mutation. Spacer collections were split, binned separately using pairwise comparisons through multithreading to lower computational time, and split bins were combined at the end. These binned spacers were mapped to loci amplicons and loci were binned according to unique patterns of spacers through multithreaded pairwise comparison. Any incomplete loci (due to trimming of amplicons) were consolidated to the largest related locus bin to compute abundance.

2.7 CRISPR immune structure

As the number of CRISPR loci in each sample varied, the dataset was rarified using the script `sinlge_rarefaction.py` script from QIIME v.1 with a subsample depth equal to the minimum number of architectures (defined in section 2.6) per sample (min=313) for all samples included in each subgroup before alpha-diversity

calculations. The expected heterozygosity index (H_{EXP}), a proxy for allele diversity, was calculated using the RAM Package v1.2.1.3 (<https://cran.r-project.org/package=RAM>).

2.8 EPV1 phage quantification

To determine the abundance of EPV1⁴⁰ in our time series dataset, we carried out targeted qPCR on DNA samples taken from our chemostat on a monthly basis during years five to eight using the following strategy: A metagenomic-based analysis was first performed to identify conserved regions within the EPV1 genome published by Skennerton et al.⁴⁰ by mapping sequenced DNA reads from days 496 and 1692 of operation against the reference phage genome. From there, we generated an optimal DNA primer set targeting 150-200 bp conserved regions using the Primer3 software and the following concentration settings: Monovalent 50mM, Divalent 1.5 mM, Oligo 50 nM, and dNTPs 0.6 nM. The selected primer set (EPV1-38198F: 5'-ACCTCGTATTCTTTGCCGGG-3', EPV1-38417R: 5'-GGCGACATATCAGGACCTGG-3') targeted the histone-like nucleoid structuring (*hns*) gene, which we also synthesized (IDT Technologies) and cloned using a TOPO TA cloning kit (Invitrogen, CA) according to the manufacturer's instructions. We also determined the optimal melting temperature by running a gradient PCR from 57-63 °C and corroborated product specificity using DNA obtained from selected *hns*-carrying colonies and samples retrieved from the chemostat.

All qPCR reactions were run in a BioRad iCycler thermocycler (BioRad Laboratories, Hercules, CA). Each reaction volume was 20 uL and contained 10 uL iQ™ SYBR®

Green Supermix (BioRAD Laboratories, Hercules, CA), 1 uL each of 10 mM forward and reverse primer, 7 mL nuclease-free water and 1 uL of sample. Templates for *hns*-based qPCR were obtained from clones carrying the synthesized *hns* gene, as described above. Ten-fold serial dilutions of the template (ranging from 10^1 to 10^7 copies per reaction) were used to generate qPCR calibration curves. The thermal cycling protocol for EPV1-*hns* quantification was as follows: initial denaturation at 95 °C for 30 s, followed by 45 cycles of denaturation at 95 °C for 30 s, annealing at °60, and extension at 72 °C for 30 s. All samples were processed in replicates, and each reaction plate contained non-template controls and standards.

Results

A laboratory-scale bioreactor was operated continuously for nine years to examine a wastewater treatment process that relies on phosphorus sequestration inside cells to remove it from the water. The bioreactor community was enriched in the genus *Candidatus Accumulibacter*, which typically constituted more than 80% of the total bacterial cells in the bioreactor, as measured using quantitative FISH ⁴⁴. The operation of this reactor and its associated microbial community has been previously described in detail ^{1,5,12,15,55,56}. Generally, two *Accumulibacter* clades (IA and IIA) dominated the microbial community, though the proportion of each clade relative to the total *Accumulibacter* abundance changed over time (Supplementary Tables 3 and 4). In earlier research, the distribution of CRISPR spacers from *Accumulibacter* clade IA against EPV1 phage in this bioreactor was found to be uneven,¹⁵ leading us to hypothesize that host-phage coevolution dynamics are actively occurring in the bioreactor. Additionally, CRISPR spacers in *Accumulibacter* clade IA did not match any of the other phage genomes previously found in this bioreactor ^{15,40}, nor *Accumulibacter* clade IIA genome CRISPR-spacers had any protospacer matching EPV1 protospacers. Therefore, the evaluation of host-phage coevolution dynamics in this manuscript, by metagenomic associations, will be specifically focused on the *Accumulibacter* clade IA / EPV1 phage pair.

Metagenomic-based evidence for differential acquisition of CRISPR spacers at the leader end

Based on clade-specific enrichment followed by shifts in the community, we selected nine samples for shotgun metagenomics sequencing (black arrows depicted in Fig. 1A). We found an uneven distribution of reads aligning to the leader-end of the array (Figure 1B). Based on this observation, we performed CRISPR-targeted amplicon DNA sequencing in 106 samples taken on a monthly basis from our bioreactors. From 2,881,846 quality-filtered DNA reads, we recovered a total of 253,737 unique CRISPR loci, different by at least one spacer. These loci contain 12,141,875 spacers in total, from which 5,224,410 spacers align to their infecting EPV1 phage (e-value < 0.01). These results show that EPV1 has been continuously interacting with its host during an extended timespan and that the susceptible host actively acquires spacers in response to phage infection events. We were unable to identify any other clear interaction between the host and other phages recovered earlier from this bioreactor,⁴⁰ although we could not rule out the possibility that a majority of the non-EPV1 spacers might correspond to either modified or acquired DNA sequences from the same virus ⁴⁰.

CRISPR loci dynamics reveal intraspecies diversity

To uncover the effect of strain-based heterogeneity in the context of CRISPR-adapted immunity, we organized the unique CRISPR loci obtained from our targeted-DNA sequencing results and plotted the top 100 most abundant ones over time. We then clustered loci based on their abundance pattern (hierarchical cluster

analysis, *hclust*, Fig. 2c). We discovered the presence of *coalitions* of CRISPR loci that coincide with the variation in relative abundance of *Accumulibacter* over time, as observed from the quantitative PCR results in Figure 2a. We define these as “related array groups” which cluster together not only by their abundance pattern but also by their spacer, as observed in the right-side panel in Fig. 2c-d.

A combined analysis of the results shown in the time series heatmap from Fig. 2 reveals an abrupt decrease in the host’s abundance followed by a significant loss in its diversity (H_{EXP}) during year six of reactor operation. This shift in the dynamics also changed the structure of the related array groups, where some of these groups were no longer present after this transition and others increased in frequency immediately after. More specifically, two cases, such as year six and mid-year seven, denote different diversity patterns as observed by a continuous decrease in the expected heterozygosity index H_{EXP} (Fig. 3a), regardless of the host’s abundance. We hypothesize that a population bottleneck driven by either phage predation and/or an operational bioreactor failure gave rise to a less complex community and such loss in diversity could potentially affect host dynamics.

To interpret how CRISPR-enabled heterogeneity could translate into enhanced immunity (or lack thereof) within *Accumulibacter* during years five and eight of operation, we traced both the EPV1 phage and host abundances and compared that with the calculated CRISPR-allele diversity, as seen in Fig. 3. These results are divided into three phases based on relative abundance and diversity: (i) both high

host's abundance and CRISPR-allele diversity; *(ii)* low host's abundance and increased CRISPR-allele diversity and *(iii)* high host's abundance and low CRISPR-allele diversity. In detail, during periods of high diversity (*i* phase) we observe high host's abundance, which was maintained until an operational crash in the chemostat occurred (loss of phosphorus-removal efficiency, supplementary Table S3). Based on EPV1's decreasing abundance, we cannot explain this fluctuation by phage predation, but instead as a result of competition within the *Accumulibacter* lineage for the same resources (dynamics of the two competing *Accumulibacter* clades depicted in Supplementary Table S4 and described in detail elsewhere¹). Next, while the host's abundance was at its lowest levels (*ii* phase), their CRISPR-loci became highly diverse until one or few host's alleles emerged in the system (middle of *iii* phase), which drives to excess viral abundance until the host's population size is reduced to a minimum at the end of the operational period, following a catastrophic community crash, wiping out most of the reactor biomass (TSS and VSS values highlighted during day 2949 of Supplementary Table S3). These contrasting observations can help explain the consequences of generating spacer diversity.

Discussion

CRISPR Diversity constitutes an important trait of the bacterial immune system, which is critical protection against viral specialization ⁵⁷. In light of the bacteria-phage arms race dynamics, a host community can develop a structure where diverse genotypes are immune to viruses because of the inherent diversity of CRISPR loci ³⁴. When triggering this response, CRISPR-active microorganisms have the potential to shift phage-host coevolution dynamics, where the host population is diverse and stable while the viral community is prone to extinction. These dynamics cannot be captured by a “kill-the-winner” model, because CRISPR-mediated diversity prevents a virus-susceptible “winner” from emerging ³⁴.

We hypothesize that a disparate spacer insertion mechanism might enable various host genotypes to prevent their extinction due to a possible phage infection event ⁵⁸, and we expect that co-existence between spacer-adding CRISPR alleles would allow them to emerge within a host population at the same time, to promote and preserve heterogeneity. Childs et al.³⁴ reported an eco-evolutionary model in which CRISPR diversity may be maintained within the host population due to clonal interference, i.e., competition among beneficially mutant clones that have independently acquired spacers in their CRISPR loci ³⁴. Different clones, each with different CRISPR spacers to the same virus, coexist and therefore prevent a sweep that would purge diversity from the bacterial community. This diversity is exclusively generated within the CRISPR locus, and thus, this mechanism does not usually provide benefits in the absence of pathogens ⁵⁷. Our results agree with this

theoretical model and experimental studies where *Pseudomonas aeruginosa* was challenged against its infecting DMS3vir phage.⁵⁹ Their results show that while the host population with low initial spacer diversity led to phage survival and escape from CRISPR immunity, phage became extinct in those populations with high spacer diversity. Although our study does not focus on EPV1-phage strategies to overcome CRISPR immunity in *Accumulibacter*, phage-bacteria coexistence experiments between *Streptococcus thermophilus* and its lytic phage 2972 reported viral genome rearrangements and escape mutations as the main strategies to promote phage persistence, supporting CRISPR as one of the fundamental drivers of phage evolution⁶⁰. We interpret these results as support for the hypothesis that CRISPR-Cas systems can readily provide their progeny with a tool to prevent future phage invasions and to enable this progeny to establish and persist in a community associated with phages. This assumption can ultimately promote an equilibrium between phage and bacteria and keep a significant amount of diversity in nature.

Eco-evolutionary implications

Coevolution can involve the development of patterns in phylogeny and in the evolution of host specificity or specialization on prey⁶¹. Interpretation of these results remains to be seen under a bacterial evolutionary scope and will constitute a study on its own. What is clear about our findings is that diversification of bacteria in response to phage predation is a possible explanation for the observed distributed immunity. The result can be a pattern that varies as the *Accumulibacter*

populations evolve and change in their distributions. CRISPR diversity promotes adaptation in dynamic environments, increasing the probability of a pre-existing genotype being fit to altered conditions ⁶². Within a community ecology context, phages hold potential to mediate coexistence among bacterial species, in which population growth by otherwise clonal bacterial species is hampered by phage infection ¹⁶. This analysis shows that viral activity might induce changes in the relative abundance of taxonomic units within an important functional group in an activated sludge community and that phage abundance, infection and immunity are previously unrecognized but important factors determining the stability of microbial communities.

Consequences and future perspectives

While the protective effect of intra-population diversity could be dramatically lost when a phage variant escapes multiple CRISPR spacers, this new phage variant is limited by the higher bacterial heterogeneity which leads to reduced population numbers of that genotype. Therefore, we propose that the additional benefit represented by CRISPR-associated distributed immunity could lead a host population to predominate under recurrent phage infection conditions under the so-called "Red Queen Hypothesis" ⁶³. The implications of this observation have a significant negative impact in biotechnological and medical applications, i.e. where phages serve as biocontrol agents in the food industry as a non-thermal intervention to reduce pathogen loads ⁶⁴ and in medicine in the form of phage

therapy, without necessarily considering the effect of bacterial strain multiplicity and their CRISPR-associated defense mechanisms. In this case, appropriate experimental evolution protocols should be often necessary for producing more effective phage preparations that could overcome the effects of CRISPR diversity⁶⁵. Likewise, greater attention must be paid to monoculture-based bacterial production at large scale, where hosts are often exposed to phage suppression. For instance, in the dairy industry, persistent phages of *S. thermophilus* present a significant threat to the production of many fermented dairy products by reducing acidification rates or even complete loss of fermentate⁶⁶. Similarly, phage activity displays a more central role in wastewater treatment systems than previously suggested^{39,67}, maintaining the microbial diversity necessary for functional redundancy in these ecosystems. Increasing the host's allele diversity would, therefore, tend to reduce the risk of phage spread, providing considerable economic consequences to the industry.

Acknowledgements

This work was partially supported by funding from the National Science Foundation (MCB-1518130 and CBET-0967646). Additional funding from the Chilean National Commission for Scientific and Technological Research (CONICYT) as a fellowship to Francisco Moya is also acknowledged. We thank the Joint Genome Institute and University of Wisconsin Biotechnology Center for DNA sequencing support, and members of the McMahon Lab for early revisions of the manuscript (Cristina Herren, Joshua Hamilton, Elizabeth McDaniel, Pamela Camejo and Ben Oyserman). The authors also thank Shaomei He, Jason Flowers and Ben Oyserman for operation of the bioreactor during the course of the study.

Figure descriptions

Figure 1 | Metagenomic-based evidence for differential acquisition of CRISPR spacers at the leader end.

(a) qPCR-based *Accumulibacter* dynamics using clade-specific *ppk1* genes. Black arrows depict nine samples selected for shotgun DNA sequencing. **(b)** Metagenomic reads aligned to the CRISPR locus belonging to the reference *Accumulibacter* clade IA UW-3 genome reveal hypervariable region (represented by a drop in coverage) within the host's locus.

Figure 2 | CRISPR loci dynamics reveal an intraspecies diversity.

(a, b), host's abundance (*ppk1* gene copies/ng DNA, **(a)** and its expected heterozygosity index (H_{EXP} , **(b)**) calculated from rarefied CRISPR-architectures obtained from the time series dataset. **(c, d)**, CRISPR-loci distribution and architectures of the top 100 most abundant alleles over time. The lateral panel represents the shape of the spacer organization within such loci. **(d)** CRISPR-related array groups and spacer identities (small subsample dataset retrieved from **(c)**).

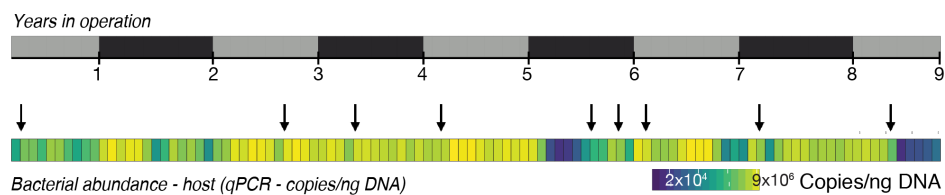
Figure 3 | Fitness consequences of generating spacer diversity.

EPV1-phage and host abundances are depicted in **(a)** and **(b)**, respectively. **(a)** The CRISPR-allele diversity (H_{EXP}) is represented by white triangles. Host-diversity phases are highlighted in shades of gray: *(i)* both increased host's abundance and CRISPR-allele diversity; *(ii)* decreased host's abundance and increased CRISPR-allele diversity and *(iii)* increased host's abundance and decreased CRISPR-allele

diversity. Operational crashes (loss of p-uptake, from supplementary table S3) are highlighted in red.

Figures

(A) Host abundance



(B) CRISPR locus architecture and metagenomic reads mapped

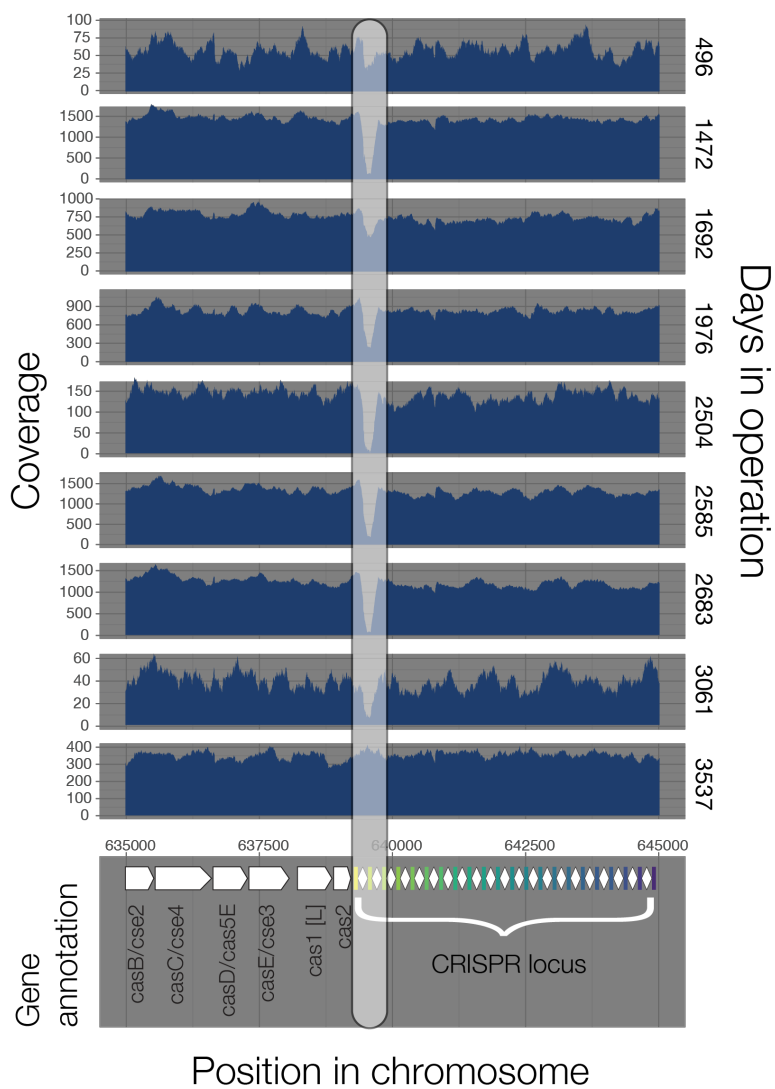


Figure 1

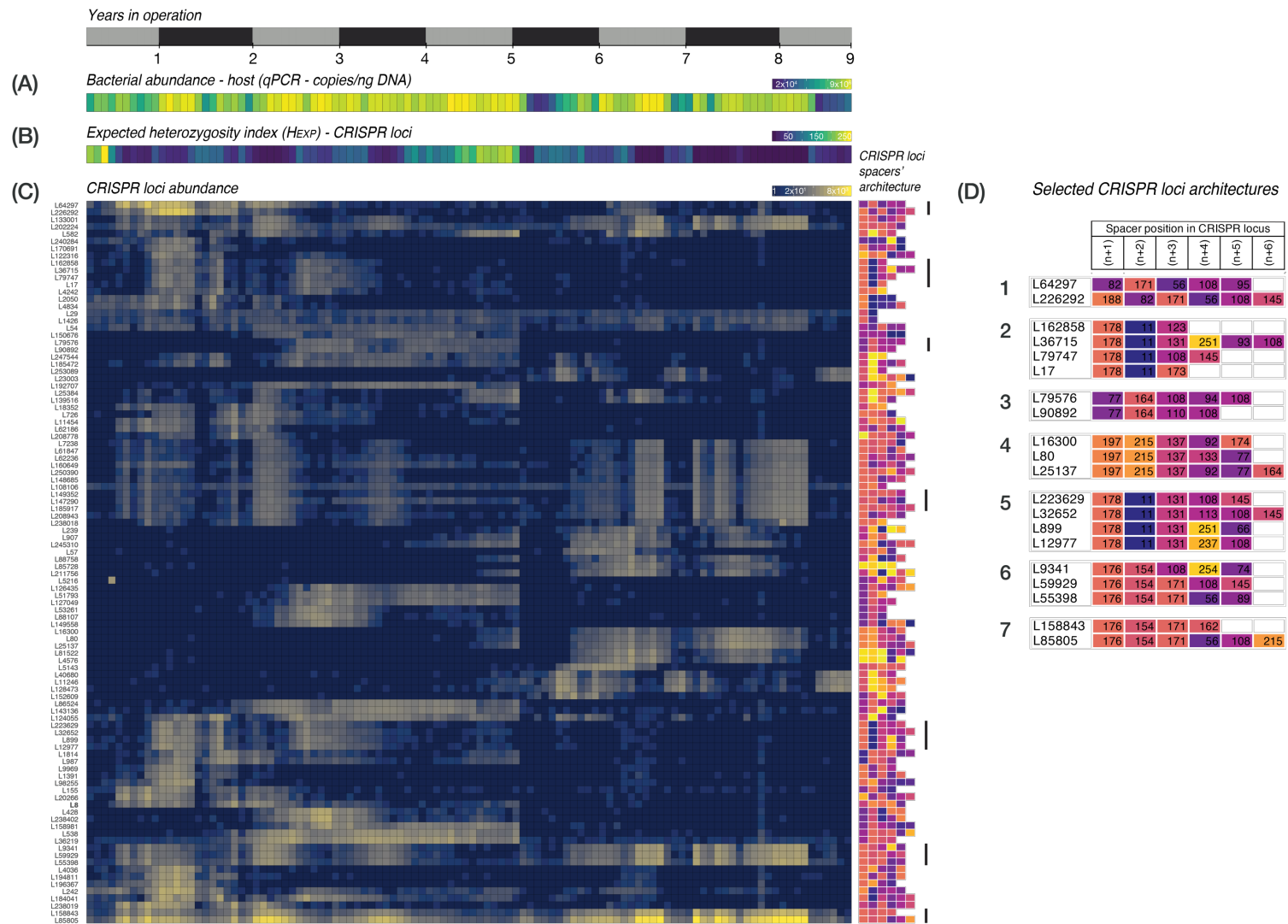


Figure 2

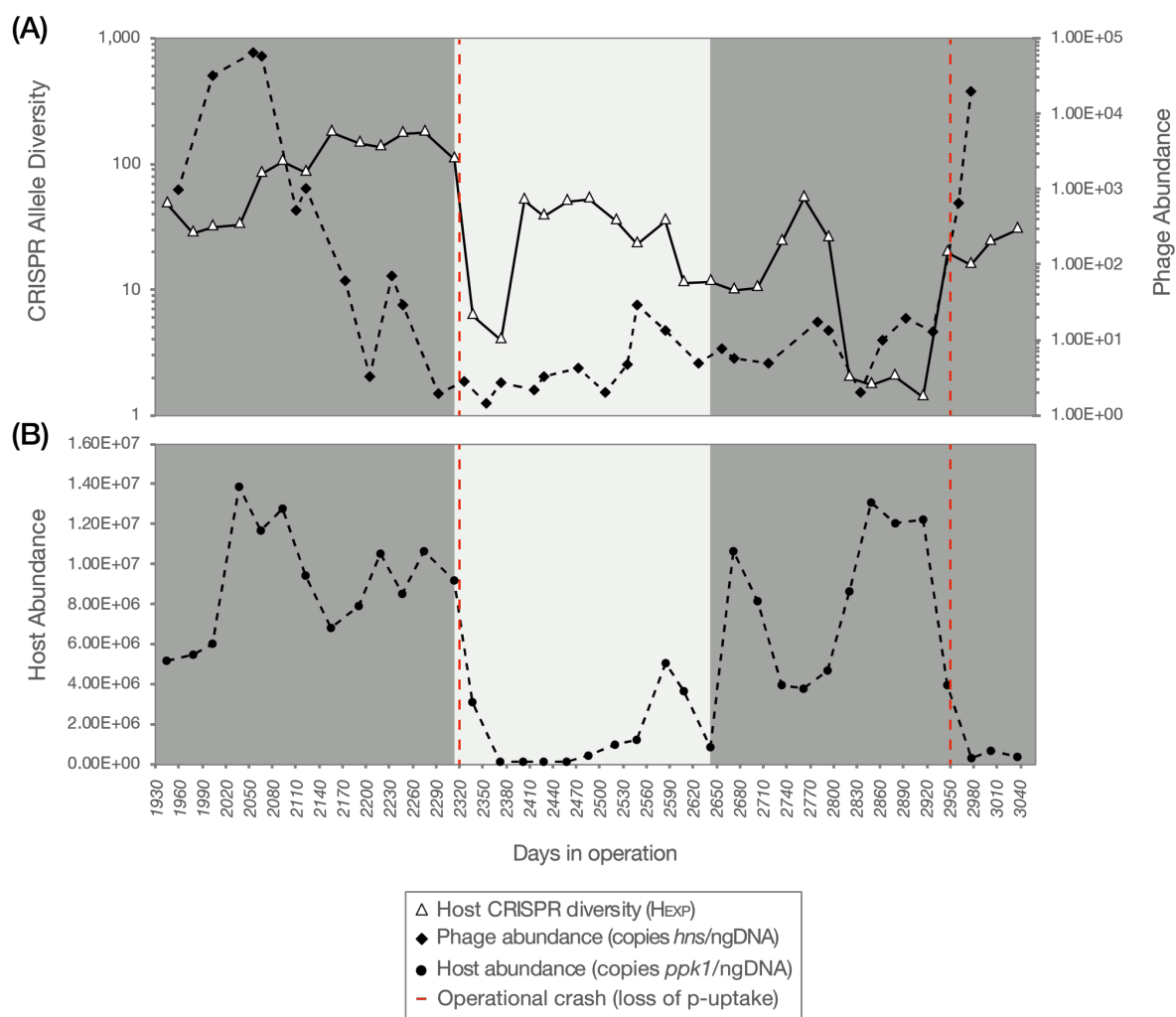


Figure 3

Supplementary Text

Samples taken when one of the two *Accumulibacter* clades dominated the PAO community in the SBR were selected for metagenomic sequencing (Supplementary table S1). Figure S1 describes heterogeneity within *Accumulibacter* clade IA-UW3 represented by ANI calculations. Metagenomic reads recruitment and ANI provide a reliable means for assessing sequence-discrete populations and determining the level of intra-population genetic diversity ⁶⁸. Typically, sequence-discrete populations share > 99% nucleotide identity to reference genomes ⁶⁹ and tend to display smaller gene-content differences among themselves. Therefore, the ANI distribution leads us to discriminate the level of heterogeneity between *Accumulibacter*-closely related populations at any given time in the dataset. From the metagenomic dataset we found that *Accumulibacter* clade IA intra-population genetic diversity accounts for less than 1% in terms of ANI distribution.

Supplementary Materials

Supplementary Figures

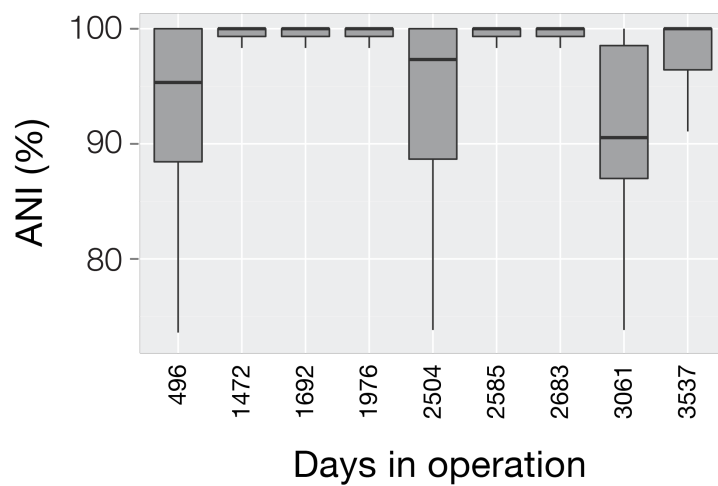


Figure S1. Average Nucleotide Identity (ANI) of metagenomic reads mapped against *Accumulibacter* clade IA reference genome.

Supplementary Tables**Table S1. Metagenomic sequencing statistics**

Metagenome sample date	Generation (SRT)	Total reads (millions)	Metagenome size (Gbp)
5/13/2013	884	97	14.5
1/23/2012	765	91	13.7
1/10/2011	671	102	15.3
10/4/2010	646	92	13.9
7/15/2010	626	89	13.3
2/2/2009	494	93	14.0
4/24/2008	423	99	14.9
9/17/2007	368	97	14.6
1/14/2005	124	100	15.0

Table S2. Reference sequences used in this study

Genome Name	Accession No.	Genome Size (†Mbp or *Kbp)	
CAP clade IA UW2	2100351004	4.5	†
CAP clade IA UW3	Ga0131788	4.3	
EBPR Podovirus 1	JF412294.1	36.1	*

Table S3. Summary of Accumulibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
9/6/03	0	0	-		-	-	-	-
6/30/04	298	75						
6/6/05	639	160	59.42	0.00	1036.00	625.33	60.36	
9/13/05	738	185						
2/9/06	887	222	58.96	0.21	1056.00	626.67	59.34	
5/20/06	987	247			1097.33	598.67	54.56	
7/17/06	1045	261						
8/3/06	1062	266	41.49	3.05	1132.00	737.33	65.14	67.30
8/10/06	1069	267	45.08	1.12	1113.33	625.33	56.17	61.11
8/21/06	1080	270						
9/19/06	1109	277	55.61	3.09	814.67	438.67	53.85	101.16
10/2/06	1122	281						
11/3/06	1154	289	55.38	0.20	937.33	537.33	57.33	129.23
11/4/06	1155	289						
11/10/06	1161	290						
11/16/06	1167	292	45.31	0.11	1090.67	612.00	56.11	
11/30/06	1181	295	58.62	0.04	1410.67	857.33	60.78	142.86
12/4/06	1185	296						
12/9/06	1190	298						
12/14/06	1195	299						
12/21/06	1202	301						
12/29/06	1210	303						
1/4/07	1216	304	26.57	0.00				
1/8/07	1220	305	26.22	0.08				
1/10/07	1222	306	38.49	0.01				
1/12/07	1224	306	36.17	0.00	650.67	436.00	67.01	
1/16/07	1228	307						
1/31/07	1243	311	25.29	0.08	577.33	270.67	46.88	
2/16/07	1259	315	60.93	1.63	1058.67	752.00	71.03	131.30
2/28/07	1271	318	55.73	0.15	1092.00	588.00	53.85	
3/12/07	1283	321						
4/24/07	1326	332	45.89	1.36	688.00	342.67	49.81	
5/1/07	1333	333	59.89	0.48	1100.00	564.00	51.27	145.33
5/14/07	1346	337						
5/18/07	1350	338						
5/22/07	1354	339	69.03	0.00				
5/24/07	1356	339	65.10	0.00	697.33	384.00	55.07	
5/29/07	1361	340	89.40	1.22	1001.33	574.67	57.39	101.98
6/1/07	1364	341	91.83	0.00	1045.33	624.00	59.69	
6/11/07	1374	344						

Table S3 (continued). Summary of Accumulibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
6/13/07	1376	344						
6/14/07	1377	344	62.90	0.55	786.67	493.33	62.71	69.78
6/17/07	1380	345						
6/21/07	1384	346	65.68	0.01	1233.33	820.00	66.49	90.42
6/27/07	1390	348						
7/10/07	1403	351	62.21	1.87	881.33	506.67	57.49	86.71
7/18/07	1411	353	40.45	0.04	745.33	510.67	68.52	52.85
8/24/07	1448	362	55.73	0.27	744.00	497.33	66.85	60.28
8/31/07	1455	364	61.51	0.00	981.33	664.00	67.66	92.07
9/7/07	1462	366	62.32	0.69	1338.67	876.00	65.44	106.52
9/14/07	1469	367	60.82	0.04	1092.00	689.33	63.13	109.00
9/23/07	1478	370						
10/1/07	1486	372	56.07	0.29	1092.00	746.67	68.38	91.66
10/8/07	1493	373	48.67	0.20	1272.00	878.67	69.08	94.96
10/15/07	1500	375	51.44	0.04	1236.00	868.00	70.23	108.59
10/22/07	1507	377	51.21	0.00	1005.33	725.33	72.15	89.60
10/29/07	1514	379	52.02	0.15	956.00	670.67	70.15	78.03
11/5/07	1521	380	46.58	0.06	1129.33	846.67	74.97	67.71
11/9/07	1525	381	57.81	0.08	989.33	700.00	70.75	71.02
11/12/07	1528	382	57.69	0.00	1145.33	836.00	72.99	117.26
11/19/07	1535	384	52.37	0.64	964.00	690.67	71.65	70.60
11/26/07	1542	386	58.27	0.08	928.00	597.33	64.37	96.20
11/30/07	1546	387	57.11	0.00	876.00	558.67	63.77	87.94
12/3/07	1549	387	39.18	0.06	534.67	332.00	62.09	58.63
12/7/07	1553	388	46.58	0.29	1065.33	585.33	54.94	161.02
12/10/07	1556	389	55.15	0.00	1034.67	602.67	58.25	151.94
12/14/07	1560	390	56.88	0.15	850.67	472.00	55.49	125.10
12/18/07	1564	391	39.06	0.06	966.67	542.67	56.14	124.28
12/20/07	1566	392						
12/26/07	1572	393	50.87	0.00	704.00	370.67	52.65	98.27
1/2/08	1579	395	52.83	0.00				
1/7/08	1584	396						
1/10/08	1587	397	60.47	0.00	821.33	486.67	59.25	76.80
1/13/08	1590	398						
1/17/08	1594	399						
1/21/08	1598	400	58.04	0.34	861.33	510.67	59.29	88.77
1/28/08	1605	401						
1/30/08	1607	402						
2/7/08	1615	404	57.00	1.66	598.67	389.33	65.03	
2/14/08	1622	406	55.61	0.43	860.00	548.00	63.72	37.57

Table S3 (continued). Summary of Accumilibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
2/21/08	1629	407	58.62	0.11	813.33	506.67	62.30	75.14
2/28/08	1636	409	49.71	0.13	800.00	534.00	66.75	60.69
3/11/08	1648	412		0.00				
3/13/08	1650	413	52.49	0.11				76.38
3/21/08	1658	415	53.64	0.08	796.00	504.00	63.32	87.12
3/27/08	1664	416	46.82	0.04	724.00	432.00	59.67	93.31
4/3/08	1671	418	49.59	0.50	726.67	474.67	65.32	74.73
4/7/08	1675	419	53.99	0.36	824.00	514.67	62.46	
4/7/08	1675	419	53.99	0.36	824.00	514.67	62.46	
4/10/08	1678	420	54.57	0.11	696.00	408.00	58.62	88.77
4/22/08	1690	423		0.00				
6/12/08	1741	435	63.71	0.15	894.67	538.67	60.21	95.79
6/15/08	1744	436		0.00				
6/19/08	1748	437	45.31	0.08	945.33	621.33	65.73	81.34
6/25/08	1754	439	58.04	0.11	878.67	572.00	65.10	85.88
7/4/08	1763	441	65.91	0.01				
7/10/08	1769	442	63.36	0.04	953.33	576.00	60.42	99.92
7/17/08	1776	444	62.99	0.00	1022.67	649.33	63.49	105.99
7/24/08	1783	446	61.12	0.00	940.00	624.00	66.38	96.01
7/31/08	1790	448	52.87	0.08	870.67	610.67	70.14	83.42
8/7/08	1797	449	48.50	0.00	914.67	569.33	62.24	
8/14/08	1804	451	61.74	0.00	908.00	564.00	62.11	
8/21/08	1811	453	54.24	0.00	980.00	597.33	60.95	99.48
8/28/08	1818	455	65.86	0.00	785.33	469.33	59.76	81.25
9/8/08	1829	457	70.73	0.00	996.00	597.33	59.97	109.90
9/15/08	1836	459	67.24	0.00	1088.00	665.33	61.15	112.07
9/22/08	1843	461	67.49	0.00	1024.00	592.00	57.81	118.14
9/29/08	1850	463	60.49	0.00	997.33	606.67	60.83	115.97
10/6/08	1857	464	67.49	0.00	988.00	596.00	60.32	111.63
10/13/08	1864	466	61.24	0.00	1013.33	606.67	59.87	111.63
10/20/08	1871	468	66.74	0.00	1082.67	620.00	57.27	127.69
10/27/08	1878	470	89.60	0.00	1221.33	740.00	60.59	134.64
11/10/08	1892	473	70.48	0.00	1072.00	665.33	62.06	123.78
11/24/08	1906	477	82.73	0.00	1008.00	633.33	62.83	121.18
12/1/08	1913	478	70.36	0.00	986.67	606.67	61.49	112.93
12/8/08	1920	480	69.86	1.23	1000.00	626.67	62.67	116.84
12/15/08	1927	482	73.98	0.00	1201.33	716.00	59.60	138.54
12/22/08	1934	484	65.86	0.00	1078.67	682.67	63.29	148.96
12/30/08	1942	486	71.11	5.05	718.67	538.67	74.95	106.42
1/19/09	1962	491	73.23	0.05	873.33	573.33	65.65	

Table S3 (continued). Summary of Accumilibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
1/26/09	1969	492	72.48	0.00	1022.67	577.33	56.45	117.27
2/2/09	1976	494	88.10	3.10	1118.67	637.33	56.97	112.07
2/5/09	1979	495	70.36	0.00				
2/9/09	1983	496	72.98	0.00				
2/13/09	1987	497	74.73	0.00				
2/16/09	1990	498	74.61	0.00				
2/20/09	1994	499	77.48	0.00				
2/23/09	1997	499	85.48	9.52				
2/26/09	2000	500	69.24	0.13				
3/2/09	2004	501	71.98	0.00				
3/5/09	2007	502	70.73	0.00				
3/9/09	2011	503	69.61	0.00				
3/12/09	2014	504	62.49	0.00				
3/16/09	2018	505	75.36	0.00				
3/19/09	2021	505	80.85	0.00				
3/23/09	2025	506	69.86	0.30				
3/26/09	2028	507	62.36	0.00				
3/30/09	2032	508	54.74	0.13				
4/2/09	2035	509	46.87	0.38				
4/6/09	2039	510	51.12	0.00				
4/9/09	2042	511	60.74	0.00				
4/13/09	2046	512	55.37	0.00				
4/16/09	2049	512	60.74	0.00				
4/20/09	2053	513	66.86	0.00				
4/23/09	2056	514	62.61	0.10				
4/27/09	2060	515	75.48	0.00				
4/30/09	2063	516	68.99	0.00				
5/4/09	2067	517	74.73	0.00				
5/7/09	2070	518	69.74	0.00				
5/11/09	2074	519	65.74	1.70				
5/15/09	2078	520	79.48	1.20				
5/18/09	2081	520	79.73	0.88				
5/21/09	2084	521	84.10	0.38				
5/26/09	2089	522	65.86	1.08				
5/28/09	2091	523	51.50	35.98				
6/1/09	2095	524	68.61	0.00				
6/4/09	2098	525	76.73	0.00				
6/9/09	2103	526	62.74	0.00				
6/11/09	2105	526	68.49	0.10				

Table S3 (continued). Summary of Accumulibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
6/16/09	2110	528	62.24	0.18				
6/18/09	2112	528	74.86	0.25				
6/22/09	2116	529	79.98	0.00				
6/25/09	2119	530	79.98	0.40				
6/29/09	2123	531	69.36	0.00				
7/1/09	2125	531	70.11	0.08				
7/9/09	2133	533	67.11	0.70				
7/13/09	2137	534	56.74	0.08				
7/16/09	2140	535	54.99	0.00				
7/20/09	2144	536	51.75	0.00				
7/23/09	2147	537	85.48	0.43				
7/27/09	2151	538	98.09	0.05				
8/3/09	2158	540	72.23	0.00				
8/6/09	2161	540	74.36	0.00				
8/10/09	2165	541	80.73	1.30				
8/13/09	2168	542	68.61	0.00				
8/17/09	2172	543	72.73	0.00				
8/20/09	2175	544	85.10	0.00				
8/24/09	2179	545	62.11	0.33				
8/27/09	2182	546	34.76	0.88				
8/31/09	2186	547	36.75	0.29				
9/3/09	2189	547	73.98	0.08				
9/8/09	2194	549	73.61	0.48				
9/10/09	2196	549	77.73	1.08				
9/14/09	2200	550	76.73	0.00				
9/17/09	2203	551	83.35	0.00				
9/21/09	2207	552	77.73	0.00				
9/24/09	2210	553	74.61	0.00				
9/28/09	2214	554	78.11	0.00				
10/1/09	2217	554	68.24	0.00				
10/5/09	2221	555	65.74	0.00				
10/8/09	2224	556	74.11	0.05				
10/12/09	2228	557	74.98	0.08				
10/15/09	2231	558	56.74	0.00				
10/19/09	2235	559	70.73	0.00				
10/22/09	2238	560	74.36	0.00				
10/26/09	2242	561	72.73	0.00				
10/30/09	2246	562	74.11	0.68				
11/2/09	2249	562	72.36	0.00				

Table S3 (continued). Summary of Accumilibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
11/5/09	2252	563	81.98	0.50				
11/9/09	2256	564	42.63	0.00				
11/13/09	2260	565	69.24	0.93				
11/16/09	2263	566	73.23	0.28				
11/20/09	2267	567	88.72	0.00				
11/24/09	2271	568	75.61	1.28				
12/1/09	2278	570	67.61	0.00				
12/3/09	2280	570	75.23	0.00				
12/7/09	2284	571	56.24	0.00				
12/10/09	2287	572	68.49	0.00				
12/14/09	2291	573	73.48	0.00				
12/18/09	2295	574	77.48	2.08				
12/21/09	2298	575	66.11	7.77				
12/28/09	2305	576	85.10	0.00				
1/4/10	2312	578	72.23	0.98				
1/11/10	2319	580	22.26	17.14				
1/14/10	2322	581	35.00	1.55				
1/18/10	2326	582	49.75	0.00				
1/21/10	2329	582	58.24	0.60				
1/25/10	2333	583	60.87	0.00				
1/28/10	2336	584	37.25	0.00				
2/1/10	2340	585	45.37	0.00				
2/4/10	2343	586	62.24	0.00				
2/8/10	2347	587	61.49	0.00				
2/11/10	2350	588	38.25	0.00				
2/15/10	2354	589	56.62	0.13				
2/18/10	2357	589	54.99	0.05				
2/22/10	2361	590	51.12	0.20				
2/25/10	2364	591	46.12	0.15				
3/2/10	2369	592	53.99	0.10				
3/4/10	2371	593	55.99	0.15				
3/9/10	2376	594	43.50	0.00				
3/12/10	2379	595	47.87	0.08				
3/16/10	2383	596	41.13	0.20				
3/18/10	2385	596	41.50	0.08				
3/23/10	2390	598	55.99	0.38				
3/25/10	2392	598	41.50	0.10				
3/30/10	2397	599	64.36	1.43				
4/6/10	2404	601	N/A	0.35				

Table S3 (continued). Summary of Accumilibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
4/8/10	2406	602	65.86	0.23				
4/13/10	2411	603	56.62	0.00				
4/15/10	2413	603	64.24	0.00				
4/20/10	2418	605	56.99	0.00				
4/22/10	2420	605	36.25	0.28				
4/27/10	2425	606	74.48	0.08				
4/29/10	2427	607	73.11	0.00				
5/4/10	2432	608	71.23	0.00				
5/6/10	2434	609	68.61	0.00				
5/11/10	2439	610	72.61	0.25				
5/13/10	2441	610	58.12	0.03				
5/20/10	2448	612	71.11	0.00				
5/25/10	2453	613	55.87	1.00				
6/1/10	2460	615	66.74	0.00				
6/3/10	2462	616	64.49	0.03				
6/8/10	2467	617	66.36	0.13				
6/10/10	2469	617	64.24	0.00				
6/15/10	2474	619	52.99	1.05				
9/7/10	2558	640	23.39	0.45	361.33	249.33	69.00	
9/13/10	2564	641	36.13	0.00	489.33	397.33	81.20	
9/20/10	2571	643	63.49	0.33	613.33	324.00	52.83	
10/1/10	2582	646	51.12	0.00	770.67	416.00	53.98	
10/11/10	2592	648	53.74	0.05	729.33	381.33	52.29	
10/22/10	2603	651	52.99	1.80	836.00	481.33	57.58	
11/8/10	2620	655	47.87	0.00	620.00	410.67	66.24	
11/16/10	2628	657	36.75	0.00	589.33	441.33	74.89	
11/23/10	2635	659	49.62	0.00	664.00	493.33	74.30	
12/3/10	2645	661	10.77	2.15	206.67	170.67	82.58	
12/16/10	2658	665	4.52	3.60	337.33	356.00	105.53	
12/23/10	2665	666	13.02	9.62	334.67	282.67	84.46	
12/31/10	2673	668	30.13	0.00	464.00	257.33	55.46	
1/7/11	2680	670	47.37	0.00	177.33	-94.67	-53.38	
1/13/11	2686	672	49.50	0.00	506.67	242.67	47.89	
2/7/11	2711	678	57.99	0.03	1209.33	762.67	63.07	
3/5/11	2737	684	73.36	1.73	966.67	602.67	62.34	
3/29/11	2761	690	60.49	2.18	832.00	452.00	54.33	
5/19/11	2812	703	66.74	0.35	486.67	278.67	57.26	
6/9/11	2833	708	64.74	0.00	1244.00	697.33	56.06	
7/5/11	2859	715	42.13	0.00	842.67	392.00	46.52	
7/29/11	2883	721	23.01	26.66	402.67	238.67	59.27	
8/4/11	2889	722	60.74	0.15	1016.00	604.00	59.45	
8/12/11	2897	724	59.24	0.13	792.00	402.67	50.84	

Table S3 (continued). Summary of Accumilibacter clade dynamics and operational parameters measured in the SBR.

Date	Oper. time (d)	Gen. (SRT)	P release (mg/L)	P uptake (mg/L)	TSS (mg/L)	VSS (mg/L)	VSS/TSS (%)	TP
8/18/11	2903	726	60.12	0.00				
8/25/11	2910	728	56.74	0.13	800.00	474.67	59.33	
9/2/11	2918	730	52.25	0.10	1081.33	572.00	52.90	
9/9/11	2925	731	51.62	0.00	716.00	396.00	55.31	
9/15/11	2931	733	73.11	0.00	940.00	530.67	56.45	
9/23/11	2939	735	75.98	0.00	950.67	557.33	58.63	
10/3/11	2949	737	23.51	6.25	393.33	321.33	81.69	
10/14/11	2960	740	49.37	0.00	684.00	474.67	69.40	
10/31/11	2977	744	61.36	0.00	732.00	458.67	62.66	
11/28/11	3005	751	25.26	4.85	628.00	525.33	83.65	
1/23/12	3061	765	54.99	0.00	1204.00	713.33	59.25	
2/17/12	3086	772	64.61	0.23	890.67	450.67	50.60	
3/12/12	3110	778	71.86	0.00	1008.00	573.33	56.88	
3/30/12	3128	782	60.74	0.00	960.00	553.33	57.64	
4/5/12	3134	784	62.24	0.00	937.33	558.67	59.60	
4/13/12	3142	786	58.62	0.00	957.33	633.33	66.16	
4/19/12	3148	787	55.49	0.00	820.00	609.33	74.31	
4/27/12	3156	789	62.24	0.00	744.00	514.67	69.18	
5/4/12	3163	791	68.11	3.70	894.67	497.33	55.59	
5/11/12	3170	793	59.62	0.10	1013.33	601.33	59.34	
5/18/12	3177	794	57.12	0.13	1277.33	776.00	60.75	
5/25/12	3184	796						
5/31/12	3190	798	64.24	0.00	1050.67	630.67	60.03	
6/28/12	3218	805	66.36	0.08	701.33	617.33	88.02	
7/30/12	3250	813	53.37	0.03	506.67	509.33	100.53	
8/27/12	3278	820	7.40	0.13	905.33	601.33	66.42	
9/20/12	3302	826	8.77	0.00	550.67	404.00	73.37	
10/8/12	3320	830	49.12	0.00	1742.67	1134.67	65.11	
11/1/12	3344	836	100.59	0.58	1046.67	998.67	95.41	
11/29/12	3372	843	51.87	0.20	465.33	402.67	86.53	
1/21/13	3425	856	58.99	0.00	756.00	465.33	61.55	
3/8/13	3471	868	52.50	0.28	722.67	452.00	62.55	
4/8/13	3502	876						
4/15/13	3509	877	59.87	0.00	685.33	430.67	62.84	
4/29/13	3523	881						
5/6/13	3530	883						
5/21/13	3545	886						
5/28/13	3552	888	0.00	6.98	556.00	550.67	99.04	

Table S4. Accumulibacter clade dynamics targeting the *ppk1* gene (Camejo, et al. (2016))

Date	Oper. time (d)	Generation (SRT)	Acc. Clade IA (copies/ng DNA)	Acc. Clade IIA (copies/ng DNA)
2/3/05	516	129	5.94E+05	4.33E+06
3/7/05	548	137	2.48E+06	1.11E+07
4/5/05	577	144	2.42E+06	5.01E+06
5/3/05	605	151	1.13E+06	8.09E+05
6/2/05	635	159	4.14E+06	6.15E+06
7/5/05	668	167	3.75E+06	9.46E+06
8/2/05	696	174	1.34E+06	4.74E+06
9/1/05	726	182	5.96E+05	1.39E+06
10/6/05	761	190	1.32E+06	2.77E+05
12/8/05	824	206	1.94E+06	1.22E+05
1/24/06	871	218	7.54E+06	1.60E+06
2/2/06	880	220	1.20E+07	2.93E+06
3/2/06	908	227	6.91E+06	5.29E+06
4/6/06	943	236	1.04E+07	2.39E+06
5/4/06	971	243	9.98E+06	2.29E+05
6/3/06	1001	250	3.74E+06	5.37E+04
6/14/06	1012	253	3.32E+05	5.96E+05
7/6/06	1034	259	9.92E+05	8.77E+05
8/3/06	1062	266	5.53E+06	1.02E+06
9/7/06	1097	274	3.00E+06	4.84E+05
10/3/06	1123	281	4.38E+06	7.84E+05
11/14/06	1165	291	2.17E+06	3.50E+05
12/5/06	1186	297	2.66E+05	6.46E+04
1/8/07	1220	305	2.88E+06	1.18E+06
2/2/07	1245	311	3.84E+06	2.01E+06
3/1/07	1272	318	9.90E+06	7.77E+05
4/4/07	1306	327	7.97E+06	3.37E+05
5/1/07	1333	333	1.20E+07	1.11E+06
6/1/07	1364	341	1.12E+07	7.66E+05
7/10/07	1403	351	8.86E+06	5.22E+05
8/1/07	1425	356	3.29E+06	1.96E+05
9/10/07	1465	366	9.63E+06	5.40E+03
10/5/07	1490	373	1.23E+07	1.35E+04
11/2/07	1518	380	1.02E+07	2.28E+04
12/7/07	1553	388	6.66E+06	9.24E+05
1/4/08	1581	395	6.68E+06	6.64E+05
2/4/08	1612	403	9.72E+06	2.43E+05
3/3/08	1640	410	7.91E+06	2.02E+06
4/1/08	1669	417	3.11E+06	1.49E+06
5/5/08	1703	426	7.90E+06	2.08E+06

Table S4 (cont.). Accumulibacter clade dynamics targeting the *ppk1* gene (Camejo, 2016)

Date	Oper. time (d)	Gen. (SRT)	Acc. Clade IA (copies/ng DNA)	Acc. Clade IIA (copies/ng DNA)
7/1/08	1760	440	7.25E+06	2.81E+06
6/2/08	1731	433	8.37E+06	2.43E+06
8/4/08	1794	449	6.49E+06	1.82E+06
9/2/08	1823	456	5.35E+06	9.20E+05
10/3/08	1854	464	8.11E+06	7.72E+05
11/7/08	1889	472	4.43E+06	5.78E+05
12/12/08	1924	481	7.22E+06	1.04E+06
1/2/09	1945	486	5.06E+06	1.03E+06
2/5/09	1979	495	5.41E+06	5.77E+05
3/2/09	2004	501	5.95E+06	9.66E+05
4/6/09	2039	510	1.38E+07	1.20E+06
5/4/09	2067	517	1.16E+07	9.24E+05
6/1/09	2095	524	1.27E+07	6.67E+05
7/1/09	2125	531	9.35E+06	5.17E+05
8/3/09	2158	540	6.70E+06	2.44E+05
9/8/09	2194	549	7.79E+06	1.31E+05
10/5/09	2221	555	1.05E+07	1.56E+05
11/2/09	2249	562	8.46E+06	1.49E+05
12/1/09	2278	570	1.06E+07	5.88E+04
1/8/10	2316	579	9.12E+06	2.51E+04
2/1/10	2340	585	2.99E+06	3.75E+06
3/9/10	2376	594	3.52E+04	8.08E+06
4/8/10	2406	602	1.75E+04	4.96E+06
5/4/10	2432	608	2.08E+04	3.49E+06
6/3/10	2462	616	5.11E+04	4.45E+06
7/1/10	2490	623	3.44E+05	7.81E+06
8/5/10	2525	631	9.06E+05	8.20E+06
9/2/10	2553	638	1.15E+06	3.49E+06
10/8/10	2589	647	4.95E+06	5.83E+05
11/1/10	2613	653	3.54E+06	3.43E+05
12/6/10	2648	662	7.36E+05	1.04E+05
1/4/11	2677	669	1.06E+07	6.45E+04
2/4/11	2708	677	8.06E+06	5.88E+04
3/8/11	2740	685	3.89E+06	2.96E+05
4/5/11	2768	692	3.65E+06	2.37E+06
5/6/11	2799	700	4.62E+06	1.47E+06
6/3/11	2827	707	8.55E+06	4.78E+05
7/1/11	2855	714	1.30E+07	2.76E+05
8/1/11	2886	722	1.20E+07	2.17E+05
9/6/11	2922	731	1.21E+07	4.02E+05
10/7/11	2953	738	3.86E+06	4.34E+05
11/7/11	2984	746	2.01E+05	4.18E+06
12/2/11	3009	752	5.75E+05	5.93E+06

Table S4 (cont.). Accumulibacter clade dynamics targeting the *ppk1* gene (Camejo, 2016)

Date	Oper. time (d)	Gen. (SRT)	Acc. Clade IA (copies/ng DNA)	Acc. Clade IIA (copies/ng DNA)
1/8/12	3046	762	7.03E+06	6.82E+05
1/6/12	3044	761	2.78E+05	1.76E+06
2/3/12	3072	768	3.75E+06	1.13E+07
3/5/12	3103	776	3.18E+05	3.68E+06
4/2/12	3131	783	3.43E+06	6.62E+06
5/4/12	3163	791	6.39E+06	5.28E+06
6/4/12	3194	799	7.34E+06	2.12E+06
7/1/12	3221	805	6.68E+06	5.74E+05
8/3/12	3254	814	4.28E+06	2.23E+06
9/3/12	3285	821	4.80E+06	3.27E+06
10/5/12	3317	829	4.79E+06	1.86E+06
11/6/12	3349	837	3.86E+06	2.92E+06
12/4/12	3377	844	5.78E+06	2.23E+06
2/8/13	3443	861	6.06E+06	1.41E+06
3/8/13	3471	868	7.31E+06	2.45E+06
4/5/13	3499	875	6.11E+06	1.43E+06
5/3/13	3527	882	5.24E+06	2.21E+06
6/3/13	3558	890	1.87E+06	1.37E+06
7/1/13	3586	897	1.62E+04	1.51E+06
9/23/13	3670	918	3.35E+04	2.09E+06
10/2/13	3679	920	3.70E+04	6.56E+05
11/1/13	3709	927	4.95E+04	1.35E+06
11/16/13	3724	931	9.59E+04	6.79E+06

Intentionally blank page

Chapter 3: Metabolic plasticity of two co-occurring Accumulibacter strains revealed through metatranscriptomics

Francisco Moya-Flores, Natalie Keene Beach, Pamela Y. Camejo, Eng Hoe Khor,
Daniel R. Noguera, and Katherine D. McMahon.

Experiment design: FMF, DRN and KDM

DNA/RNA processing and library preparation: FMF and EHK

qPCR analysis: NKB

Data analysis: FMF, PYC and MK

Manuscript writing: FMF and KDM

Abstract

The activated sludge process has been recently used as a model ecosystem to understand microbial interactions and performance in both laboratory and full-scale systems. This microbial community is enriched in the uncultured Polyphosphate Accumulating Organism (PAO) *Candidatus* Accumulibacter phosphatis. Accumulibacter constantly adapts its global physiological response across biphasic cycles of feast and famine conditions by simultaneously cycling multiple intracellular biopolymers. Storage polymer cycling ultimately allows Accumulibacter to sequester carbon under anaerobic conditions for use in subsequent aerobic conditions, resulting in its numerically and functionally significant abundance in the EBPR process. Recent investigations have suggested that Accumulibacter clades are niche-specific and their differences could offer a competitive advantage for specific clades in a mixed community, depending on the operating conditions. However, no study has shown differences in transcriptional profiles amongst two or more Accumulibacter clades, and never before in the clade IA strain. Here we use comparative metatranscriptomics to characterize the molecular response of two co-existing Accumulibacter clades exposed to the same environmental stimuli. A combination of metagenomic and RNA sequencing was used to identify differentially expressed genes between Accumulibacter clade IA and IIC across the EBPR cycle. Our study shows that Accumulibacter's inter-clade differentiation may have important implications for process performance.

Main

Activated sludge wastewater treatment processes are ubiquitous for the removal of organic matter and nutrients from municipal and industrial wastewater. These systems employ complex microbial communities to efficiently remediate wastewater streams prior to being discharged into the environment. When phosphorus removal is the desired outcome of activated sludge wastewater treatment, Enhanced Biological Phosphorus Removal (EBPR) is implemented by selectively enriching Polyphosphate Accumulating Organisms (PAO) in alternating anaerobic and aerobic conditions ^{4,70,71}. These alternating conditions create feast-famine cycles that seem to promote biopolymer cycling for subsequent utilization, probably associated to a stress-response mechanism. Under anaerobic conditions, PAOs take up volatile fatty acids (VFAs) and convert them into polyhydroxyalkanoates (PHA). PHA production requires energy (ATP) and reducing power. Current metabolic models assume that ATP is supplied by polyphosphate degradation and, to a lesser degree, glycogen degradation ^{5,70,72–74}. Reducing power is provided by glycogen degradation and the TCA cycle. In the aerobic phase, when oxygen is available for respiration, VFAs are not present in the medium for other species and the PHA reserves of PAOs ensure their dominance in the microbial ecosystem. The restoration of polyphosphate reserves via ATP, depletes the water of ortho-Phosphate, thus achieving treatment goals ^{4,5}.

In most EBPR processes, the dominant PAO is a member of the Betaproteobacteria in the Rhodocyclus group, named *Candidatus Accumulibacter phosphatis* ^{70,75}.

Accumulibacter is subdivided into two main types (I and II). These types may be further subdivided into several clades which demonstratively have different ecotypes and thus inhabit different niches within the EBPR ecosystems^{8,10-12}. Several researchers have investigated kinetics and substrate specificity of Accumulibacter clades, which appear to have ecologically distinct characteristics. Preliminary research indicates that Accumulibacter clade IA has higher acetate uptake rates and higher phosphate release rates, and that this clade can reduce nitrate while Accumulibacter clade IIA cannot^{12,76}. It has also been shown that Accumulibacter Type II is able to switch to partial glycogen degradation more quickly than Type I, thus enabling Type II cells to fuel VFA uptake when polyphosphate (poly-P) becomes limiting⁷⁷. Welles et al. demonstrated that the two Accumulibacter clades may use different metabolic pathways for anaerobic conversions, with phosphate availability being a key factor affecting their competition and thus influencing the anaerobic stoichiometry⁷⁸. Specifically, VFA uptake rates for Accumulibacter type I are greater than of type II when the concentration of polyphosphate in the bulk liquid is high. The opposite occurs when polyphosphate concentration is low. This seems to be due to differences in the two clades' ability to shift between a polyphosphate-accumulating and a glycogen-accumulating metabolism. These results suggest that Accumulibacter clades are niche-specific and their differences could offer a competitive advantage for specific clades in a mixed community, depending on the operating conditions. This inter-clade differentiation may have important implications for process performance.

New bioinformatics tools have enabled researchers to search for explanations to population abundance and bioreactor performance variability at the molecular level⁵⁵. For instance, Flowers et al. compared the gene content between strains of *Accumulibacter* Clades IA and IIA enriched in a lab-scale bioreactor¹⁵. Although core genome-level differences were not detected between the two clades, *Accumulibacter* Clade IIA had the potential to fix carbon and nitrogen and contained an overrepresentation of a protein complex which may allow for it to operate either a reductive or complete TCA cycle anaerobically¹⁵. Another comparative genomic analysis revealed a substantial number of clade-specific genes, where the main differences across those clades were the type of nitrate reductase encoded and the capacity to perform subsequent steps in denitrification²⁵. Similarly, the advent of next-generation sequencing has enabled numerous transcriptional investigations focused on various aspects of *Accumulibacter* metabolism^{55,79–81}. A time series metatranscriptome generated from enrichment cultures of *Accumulibacter* clade IIA was used to gain insight into anaerobic/aerobic metabolism and regulatory mechanisms within an EBPR cycle⁴⁴. Co-expressed gene clusters were identified, displaying ecologically relevant trends consistent with batch cycle phases. Some of the debatable features of *Accumulibacter*'s anaerobic operation were confirmed under these approaches, alleviating most of the questions that wastewater researchers have explored over the past two decades. However, due to the phenotypic differences discussed above, a characterization of *Accumulibacter*'s immediate response to the shifts in oxygen availability, at the lineage level, remains to be discovered. Likewise, no study has shown differences in transcriptional

profiles amongst two or more *Accumulibacter* clades, and never before in the clade IA strain. Here, we use comparative metatranscriptomics to characterize the molecular response of two co-existing *Accumulibacter* clades, enriched in a lab-scale bioreactor performing P-removal. A combination of metagenomic and RNA sequencing was used to identify differentially expressed genes between *Accumulibacter* clade IA and IIC across the EBPR cycle.

Methods

3.1 Bioreactor operation

A laboratory-scale sequence batch reactor (SBR) was seeded with activated sludge from the Madison Metropolitan Sewerage District (Madison WI, USA) during August 2015, and the enrichment was operated as described by Garcia-Martin et al.³. The bioreactor community was enriched in the genus *Accumulibacter*, which typically constituted more than 80% of the total bacterial cells in the reactor, as measured using Fluorescence in situ Hybridization (data not shown). The biphasic feast-famine cycle of the SBR is depicted in Supplementary Figure 1, as the average profile of the reactor during two consecutive years of operation. The SBR was operated with the established biphasic feast-famine conditions: a 6 h cycle consisted of the anaerobic phase, 110 min; aerobic, 180 min; settle, 30 min; draw and feed with an acetate-containing synthetic wastewater feed solution, 40 min. A reactor hydraulic residence time (HRT) of 12 h was maintained by withdrawing 50% of the reactor contents after the settling phase, then filling the reactor with fresh nutrient feed; a mean Solids Retention Time (SRT) of 4 d was maintained by manually withdrawing 25% of the mixed reactor contents each day immediately prior to a settling phase.

3.2 Experimental design

Samples for gene expression analyses were collected at different time points during typical transitions across a single cycle, following the same sampling strategy

followed by Oyserman et al.⁴⁴: (1) Anaerobic, no carbon contact, (2) Anaerobic, early carbon contact, (3) Anaerobic, late carbon contact, (4) Early aerobic, high phosphorus, (5) Mid-aerobic, (6) Late aerobic, low phosphorus. These samples were flash-frozen on liquid nitrogen immediately after centrifuging and discarding the supernatant. Standard measures of phosphorus and carbon dynamics were performed at the same times.

3.3 DNA and RNA extractions

Samples were subject to TRIzol-based RNA extraction (Thermo Fisher Scientific, Waltham, MA) followed by phenol-chloroform separation and RNA precipitation. RNA was purified following an on-column DNase digestion using the RNase-Free DNase Set (Qiagen, Venlo, Netherlands) and cleaned up with the RNeasy Mini Kit (Qiagen, Venlo, Netherlands). A phenol-chloroform bead-beating extraction method was performed to obtain total genomic DNA from the biomass in preparation for shotgun metagenomic sequencing (details in section 2.1).

3.4 *Accumulibacter* clades quantification

To determine the relative proportions of the enriched *Accumulibacter* clades within the bioreactor microbial community, we used quantitative Polymerase Chain Reaction (qPCR) using clade-specific primers targeting the polyphosphate kinase (*ppk1*) gene as described Camejo et al.⁴⁵.

3.5 Library preparation and DNA sequencing

Metagenomic sequencing of the SBR was performed in collaboration with the UW-Biotechnology Center (three samples per Illumina HiSeq lane). Briefly, microbial DNA was size-selected by targeting 550 bp long products. The products were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA Biosystems). The libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche Light Cycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq2500 sequencer using TruSeq SBS sequencing kits, following a 2x150 indexed run recipe. Raw data for this study consisted of 219.4 million 300-bp Illumina HiSeq reads with about 3.9 Gpb per sample (Table 1 – supplementary material).

3.6 Construction of Prokaryotic Illumina RNA Libraries

Total RNA submitted to the University of Wisconsin-Madison Biotechnology Center was verified for purity and integrity using a NanoDrop 2000 Spectrophotometer and an Agilent 2100 BioAnalyzer, respectively. RNA-Seq paired end libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA). Each sample was processed for ribosomal depletion, using the Ribo-Zero™ rRNA Removal kit (Bacteria). mRNA was purified from total RNA using paramagnetic

beads (Agencourt RNAClean XP Beads, Beckman Coulter Inc., Brea, CA) and fragmented by heating in the presence of a divalent cation. The fragmented RNA was then converted to cDNA with reverse transcriptase using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, California, USA) with random hexamer priming and the resultant double stranded cDNA was purified. cDNA ends were repaired, adenylated at the 3' ends, and then ligated to Illumina adapter sequences. Quality and quantity of the DNA were assessed using an Agilent DNA 1000 series chip assay and Thermo Fisher Qubit dsDNA HS Assay Kit. Libraries were diluted to 2nM, pooled in an equimolar ratio, and sequenced on an Illumina HiSeq 2500, using a single lane of paired-end, 100 bp sequencing, and v2 Rapid SBS chemistry. FASTQ files were generated using Casava 1.8.2.

3.7 Raw read processing and de novo assembly of metagenomic reads

Illumina unmerged reads were quality filtered and trimmed using the Sickle software v1.33⁸². Reads were merged with FLASH v1.0.3²²⁸³, with a mismatch value of ≤ 0.25 and a minimum of ten overlapping bases from paired sequences, resulting in merged read lengths of 150 to 290 bp. FASTQ files were then converted to FASTA format using the Seqtk software v1.0⁸⁴. Metagenomic reads were then co-assembled using the Velvet assembler⁴⁷ with a k-mer size of 65 bp, a minimum contig length of 200 bp, and a paired-end insert size of 300 bp. Metavelvet⁸⁵ was used to improve the assembly generated by Velvet.

Genome binning from the metagenomic assembly was performed using Maxbin⁴⁹.

Genome completeness and redundancy was estimated using

CheckM ⁵⁰. Contigs were scaffolded using Medusa ⁸⁶, manually inspected for short contigs removal and decontaminated using the JGI tool Prodege ⁵² and Anvi'o ⁸⁷. Further scaffolding was performed on the MAGs using Nanopore long-reads and the LINKS algorithm ⁸⁸. [Gapcloser](#) was used for additional gap filling.

3.8 Genome annotation and metabolic reconstruction

Metabolic reconstruction of the recovered draft genomes was performed using the JGI pipeline ⁸⁹. Annotated genomes were then used to reconstruct the metabolic network of each organism using Pathway Tools ⁹⁰ and the MetaCyc database ⁹¹. Pathway/Genome Databases were created for each genome, where pathway inference was based on a set of rules used by the Pathway Tools prediction algorithm Pathologic, including the presence of all key reactions and the completeness of the reconstructed pathway. All inferred pathways were then manually curated to verify predictions.

3.9 Identification of orthologous genes

We used reciprocal best BLAST hits ⁹² to identify a set of core genes representing the *Accumulibacter* pangenome. Sequence information for *Accumulibacter* spp. IA-UW3 and IIC-UW6 were based on the Flowers et al. ¹⁵ and our assemblies, respectively. In the case of multiple gene copies matching to a single ortholog, we used the one with the highest percent identity, resulting in 1,294 one-to-one ortholog pairs.

3.10 Definition of functional gene sets

Functional KEGG categories were downloaded from <https://www.genome.jp/kegg/ko.html>. Assignment of genes from *Accumulibacter* to KO terms was automatically performed by KEGG. We developed an in-house R script to extend the KO terms to include parent levels (category, sub-category, and ontology). All subcategories that contained at least ten ontologies were considered for functional analyses.

3.11 Metatranscriptomics analysis

RNA reads were quality filtered and trimmed with Sickle and forward and reverse reads were merged using FLASH (v. 1.2.11) ⁸³. Ribosomal RNA sequences were removed with SortMeRNA using six built in databases for bacterial, archaeal and eukaryotic small and large subunits ⁹³. Resulting non-rRNA reads were mapped to all assembled metagenomic contigs using BBSplit, a tool for mapping to multiple references simultaneously, from BMap (v35.92) ⁹⁴, with minimal identity of 95%. Read counts were calculated for each predicted ORF using htseq-count v0.6.0 and normalized by sequencing depth and expressed as transcript per million (TPM) values using a custom R script ⁸¹.

3.12 Differential gene expression calculations

Identification of significant changes in transcript abundance was carried out using the RNAentropy software ⁹⁵. This program identifies genes or transcripts with a significant variation of expression across all conditions studied, by performing a

statistical tests that assess if and how much the expression of a gene across any number of different samples diverges from a given background.

Results

Metagenomic sequencing and binning of Accumulibacter genomes

Metagenomic sequencing of SBR biomass was performed in collaboration with the UW-Biotechnology Center. We assembled four highly complete draft genomes initially classified in the *Rhodocyclaceae* family based on phylogenetic analyses of specific marker genes⁸. The estimated completeness and contamination of these genomes (Table 1) suggest that three correspond to high quality metagenome-assembled genomes (MAGs), and one to a medium quality MAG⁹⁶. Neighbor-joining phylogenetic tree analysis of the *ppk1* gene revealed that these genomes belonged to Accumulibacter clades IA, IIA, IIC, and IIF (data not shown). Average Nucleotide Identity (ANI) comparisons between these clades and a collection of existing Accumulibacter genomes revealed above 90% ANI and >68% of aligned fraction with other Accumulibacter genomes, confirming species identification within the Accumulibacter lineage (Supplementary Figure S2). Remarkably, MAGs IA-UW4 and IIA-UW5 assembled in this study had a high ANI of 99.3% and 99.8%, respectively (alignment fractions of 0.768 0.933, respectively), with genomes of these clades assembled ten years ago from the same source of the reactors' inoculum, suggesting little temporal divergence of their genotypes.

All four MAGs were submitted to the Joint Genome Institute for automatic Open Reading Frame (ORF) prediction and gene annotation using their standard pipelines⁸⁹. We were unable to obtain a high-quality draft genome for Accumulibacter clade IA-UW4 (85.01% completeness and 3.41% redundancy), possibly due to a high

strain heterogeneity within such clade. As a result, we used the *Accumulibacter* IA-UW3 genome ¹⁵ as a reference for our analysis, since it shared the highest ANI (99.67%) with the recovered MAG. Metabolic pathways were then curated for each of the MAGs, using Pathways tools ⁹⁰ and compared against JGI's predictions.

Metatranscriptomics profiling of *Accumulibacter* clades under classical EBPR conditions

RNA sequencing from seven samples extracted over the course of the EBPR cycle yielded a total of 186.1 million reads, after quality filtering (Supplementary Table S1). Approximately, 55% of those reads corresponded to mRNA and 8.6% and 27.6% of them mapped to the *Accumulibacter* clade IA-UW3 and IIC-UW6 reference genomes, respectively (Supplementary Table S3). This proportion of alignment indicates that, overall, *Accumulibacter* accounts for a considerable amount of the transcriptomic activity in the chemostat. The other two *Accumulibacter* strains present in the reactor (*Accumulibacter* IIA-UW5 and *Accumulibacter* IIF-UW7) only retrieved 1.02% of the total mRNA reads, and therefore, were not included in the following analysis.

Identification of genes showing a significant variation of expression within the seven RNA-Seq samples was done using entropy calculations ⁹⁵. Transcripts showing a significant variation (local p-value > 0.01) across all time points studied were considered to be dynamic, or differentially expressed (DE). The expression patterns of each gene were stored in a trinary logic table, where each gene was assigned a -1, 0, or 1 for each time point, where -1 corresponds to gene downregulation, 0 to

basal expression, and 1 to upregulation. This data was subsequently used to compare transcription between both genomes.

To compare broad patterns in DE, we performed a reciprocal BLASTP analysis of genomes *Accumulibacter* IA-UW3 and IIC-UW6. We identified 2,376 orthologous genes present in both genomes (core genome), and 1,477 and 2,522 genes unique to *Accumulibacter* IA-UW3 and IIC-UW6, respectively (flexible genome). Combining this orthologous analysis with transcriptional data, we constructed a double-layered Venn diagram depicting *Accumulibacter*'s pan-transcriptome to evidence unique and common transcriptomic traits associated to the same condition (Figure 1 and Supplementary Table S2). Several categories were identified using this strategy:

- **Core-DE:** Set of core orthologous genes that were differentially expressed in both *Accumulibacter* genomes.
- **Core, DE, one *Accumulibacter* Clade only:** Set of core orthologous genes that were differentially expressed in only one of the two *Accumulibacter* genomes.
- **Flexible genome, DE:** Genes unique to one of the two genomes that were differentially expressed.

We found an overrepresentation of core genes that were differentially expressed only in *Accumulibacter* Clade IA (46%) versus IIC (33.8%). The latter's flexible genome accounted for 25.6% of the differentially expressed genes, and the former for only 17%; which might be the results of both genome streamlining and niche-specialization.

To determine functional differences in *Accumulibacter*'s gene expression, orthologous genes were classified using the KEGG database (Figure 2). Using this approach, in *Accumulibacter* Clade IA-UWC, we identified an overrepresentation of DE genes corresponding to the pathways Oxidative phosphorylation (PATH:ko00190), Glyoxylate and dicarboxylate metabolism (PATH:ko00630), Glycolysis / Gluconeogenesis (PATH:ko00010), Pyruvate metabolism (PATH:ko00620), Starch and sucrose metabolism (PATH:ko00500), Transcription factors (BR:ko03000), Protein folding and associated processing and Folate biosynthesis (PATH:ko00790). Most of these pathways belong to carbohydrate metabolism, indicating a clear preference of this clade towards substrate conversion and utilization. On the other hand, the majority of *Accumulibacter* Clade IIC-UW6 DE genes, were classified as part of: Purine metabolism (PATH:Ko00230), Ribosome biogenesis (BR:ko03009), Energy metabolism, two-component system (PATH:ko02020), Porphyrin and chlorophyll metabolism (PATH:ko00860), Protein kinases (BR:ko01001) and Amino sugar and nucleotide sugar metabolism (PATH:ko00520).

Acetate conversion/transport and PHB synthesis

In *Accumulibacter*, acetate is first taken up by active transport and activated to acetyl-CoA by low and high-affinity routes. Two acetyl-CoA molecules are then condensed into acetoacetyl-CoA and subsequently reduced to poly- β -hydroxybutyrate^{97,98}. We examined the expression patterns of key genes involved in PHB synthesis: phosphate acetyltransferase (*pta*), acetate kinase (*ack*), and

acetyl coenzyme A synthetase (*acs*). We also examined the polyhydroxyalkanoate (*pha*) genes in the *Accumulibacter* genomes which are involved in polyhydroxybutyrate generation. Interestingly, both clades repressed the expression of the high-affinity *acs* gene, while favoring the low-affinity, acetyl-phosphate producing, *ack/pta* route (Figure 3). These results differ with previous observations^{81,97}, where *Accumulibacter* clades IIA-UW1 and IC_{LDO} exhibited upregulation of genes associated to both low and high-affinity routes. We attribute this discrepancy to their slow-feeding strategy (60 min vs. 20 minutes of acetate addition in Oyserman et al. 2016 and Camejo et al. 2018 studies), which might favor acetate assimilation via the high-affinity route when cells are scavenging for small amounts of environmental acetate.

Next, in both genomes, we identified several of genes encoding PHA synthesis: *phaA*, *phaB*, and *phaC* with two, four and three paralogs, respectively. Additionally, we detected single copies of *phaE* and *phaR* in both *Accumulibacter* genomes. Except for one *phaC* copy in each genome, all *phaABC* genes were upregulated early upon acetate contact, with instant repression in the middle of the acetate feeding phase, potentially indicating a rapid substrate concentration-dependent gene activation in *Accumulibacter*. Notably, such repressed *phaC* gene followed the same repression pattern of a *phaE* gene. To understand such inconsistencies in transcription, we investigated the nature of all *phaC* genes by running a conserved protein domain analysis using the CDD/SPARCLE tool⁹⁹. This analysis revealed that these three paralogs encode two Class I and one III, enzymes with differential subunit composition and substrate specificity. Class III PHA synthase requires two

different types of subunits (*phaE* and *phaC*) and Class I is a single subunit-containing enzyme ¹⁰⁰. Oyserman et al.⁹⁷ reported the same results for *Accumulibacter* Clade IIA-UW1. These observations are intriguing since, to date, no other bacterium containing two different PHA synthase classes has been identified, suggesting not only a redundant mechanism in *Accumulibacter* but also a versatile one.

Anaerobic TCA operation

In *Accumulibacter* IA-UW3, most TCA-related genes (except for 2-oxoglutarate dehydrogenase) were upregulated during early acetate contact and repressed before the end of the anaerobic phase. Instead, the activity of 2-oxoglutarate dehydrogenase was replaced by a 2-oxoglutarate:ferredoxin oxidoreductase (KFOR), in which expression was basal. In contrast, *Accumulibacter* IIC-UW6 repressed KFOR immediately after acetate addition. In turn, the classical 2-oxoglutarate dehydrogenase gene was early upregulated in the anaerobic phase and subsequently downregulated aerobically. Furthermore, from the two citrate synthase (*cs*) copies encoded by *Accumulibacter* IIC-UW6, one was repressed after acetate contact whereas the other showed no change in expression during that stage. This may be due to a high concentration of reducing equivalents in the form of NAD(P)H that inhibit *cs* to ensure the availability of acetyl-CoA for *phaB* activity

¹⁰¹.

We further evaluated the activity of the glyoxylate cycle (Figure 3) and observed upregulation of the malate synthase and isocitrate lyase genes in *Accumulibacter*

IA-UW3 during the anaerobic phase, with subsequent aerobic repression. Contrastingly, we did not find any evidence of the anaerobic operation of the glyoxylate cycle in *Accumulibacter* IIC-UW6, where malate synthase was repressed, and isocitrate lyase did not show any differential expression across the entire cycle.

Respiration

To understand the response of aerobic respiration pathways to shifts in oxygen availability, we examined four protein complexes related to aerobic respiration: the respiratory chain supercomplex I-III-IV (formed by NADH dehydrogenase, Ubiquinol-cytochrome C reductase and low (*aa3*) and high (*cbb3*) -affinity cytochrome C oxidases) and complex II (Fumarate reductase / Succinate dehydrogenase). For both *Accumulibacter* genomes, respiratory complexes I and III were upregulated upon acetate contact during the anaerobic phase in a similar fashion. Complex IV, represented by the terminal oxidases *aa3* and *cbb3*, exhibited distinct transcriptional profiles, where the high-affinity cytochrome oxidase was upregulated after acetate contact in the anaerobic phase and the *aa3*-type cytochrome oxidase was only expressed after the redox transition, suggesting that these two complexes might undergo transcriptional regulation based on oxygen availability ⁸¹.

Similarly, two succinate dehydrogenase/fumarate reductase complexes exhibited divergent expression patterns when carbon or oxygen was fed in the chemostat. Further analysis of the conserved domains of these gene clusters ⁹⁹, revealed that they corresponded to the QFR (quinol:fumarate reductase) and SQR

(succinate:quinone reductase) families. QFR is known to participate in anaerobic respiration with fumarate as a terminal electron acceptor, and SQR is involved in aerobic metabolism as part of the TCA cycle and the aerobic chain. Surprisingly, SQR was upregulated during anaerobic acetate contact, suggesting anaerobic TCA activity, whereas QFR was repressed during the later anaerobic stage. We further explored a previous *Accumulibacter* UW1 transcriptomics dataset ⁹⁷, in which we observed the same trends for SQR and QFR activity, i.e., anaerobic repression of the fumarate reductase complex and upregulation of the succinate dehydrogenase upon anaerobic acetate contact. It has been hypothesized that electrons from FADH₂ generated by SQR are transferred to NAD(P) by the novel cytochrome b/b6 (cyb/b6), which consists on a fusion of one gene with a cytochrome b/b6 domain and another gene with soluble ferredoxin, NAD(P)- and flavin-binding domains ^{15,102,103}. We identified paralogs of such genes in both *Accumulibacter* genomes; however, only the one found in *Accumulibacter* clade IA was anaerobically expressed. In fact, upon anaerobic acetate contact, the copy encoded by *Accumulibacter* clade IIC was repressed. Interestingly, we did not find a flavin-binding domain in this gene, as opposed to *Accumulibacter* IA, suggesting an alternative to a full anaerobic TCA cycle in this genome.

Denitrification capabilities

We scrutinized the potential of both genomes to fully at the genome and transcriptomic levels. Results of differential expression throughout the cycle are depicted in Figure 4. In agreement with previous observations ²⁵ for other clade IIC

genomes, *Accumulibacter* IIC-UW6 encodes a respiratory nitrate reductase (*nar*), a trait that could potentially enable its dominance under anoxic conditions. Our results indicate of the *narGHJ* operon during the anaerobic phase. Conversely, a nitrite reductase (*nirS*) gene was also present in the genome and upregulated during early anaerobic. On the other hand, *Accumulibacter* IA-UW3 does encode for most of the genes required for anaerobic respiration when nitrate is available^{15,25,81}. A periplasmic nitrate reductase *napADFGH* operon, a ferredoxin-type *napGH* operon and a cytochrome c-type *napC* gene were all upregulated upon anaerobic acetate contact, probably due to transcriptional regulation as nitrate was not present during that phase. It is noteworthy to mention that the gene *napF* was not previously identified in the *Accumulibacter* IA-UW3 genome¹⁵, and we provide evidence of its presence, probably due to updated gene annotation tools. Likewise, one copy of a nitrite reductase (*nirS*) and nitrous oxide reductase (*nosZ*) genes were upregulated anaerobically when carbon was available. Camejo, et al. reported the absence of a nitric oxide reductase, q-NOR-like (*norZ*) gene in this genome⁸¹, suggesting that either *Accumulibacter* IA-UW3 encodes for an alternative route to nitric oxide reduction or an incomplete denitrification pathway. Instead, we identified cNOR (cytochrome c dependent nitric oxide reductase) subunits D and Q present in this genome and upregulated during the late aerobic phase.

Discussion

Using a combination of metagenomics and metatranscriptomics, we retrieved several high-quality *Accumulibacter* genomes from a laboratory-scale EBPR bioreactor, accounting for 37% of the total mRNA reads in the chemostat (Supplementary table S3). Our work presents the first study to understand both the genomic and transcriptomics differences of two simultaneously-enriched *Accumulibacter* strains (spp. IA and IIC), which are thought to occupy different niches in the activated sludge process^{12,81,98}. Remarkably, the *Accumulibacter* clade IA MAG recovered from our chemostat was almost identical to the *Accumulibacter* IA-UW3 genome published by Flowers et al.¹⁵, suggesting that its ecotype can indeed be maintained throughout space and time.

Our analysis framework included normalization of transcripts per million data using htseq and the identification of differentially regulated genes using a statistical significance test based on RNentropy calculations. The amplitude of changes in mRNA levels was disregarded, largely because such changes are subject to the magnitude of the perturbations both in intensity and duration. We discretized the regulation of differentially expressed genes into three levels; upregulated (+1), downregulated (-1) and constitutively expressed (0), to avoid biases by the contribution of individual experiments. We interpret an increase in the mRNA levels to imply the operation of a related biochemical pathway at that time, acknowledging that genes can be up-or down-regulated in response to specific environmental signals, making a protein or pathway available for a subsequent phase. We also

acknowledge that anaerobic upregulation does not necessarily mean that the protein is essential for anaerobic EBPR activity. However, we follow the most parsimonious interpretation, i.e., expression causes activity in that phase ⁵⁶.

Further examination of the *Accumulibacter* transcriptome shows that a large number of genes were differentially expressed under the regular EBPR conditions. We identified 1,294 genes that exhibit consistent differential regulation in both *Accumulibacter* strains, and we designate these genes collectively as the core transcriptional response (CRT) in *Accumulibacter*. However, the only group of genes that had an statistically significant correlation at the CRT level corresponded to those belonging to KEGG orthologies bacterial motility proteins and oxidative phosphorylation (Pearson product-moment correlation coefficient at $p > 0.1$). We found most of these orthologous grouped in operons across both *Accumulibacter* genomes (data not shown), which explains a regulatory response to the conditions studied.

Clade-specific responses to anaerobic carbon uptake

A central feature of *Accumulibacter* physiology is the rapid uptake and sequestration of acetate in the form of PHA. Past studies have demonstrated that NAD(P)H production via glycogen degradation is not sufficient to explain the observed levels of PHA in acetate-fed systems ⁷⁰. It is currently accepted that different versions of the TCA cycle operate in the anaerobic phase to provide the extra reducing power needed to balance PHA storage and PAO survival across variable redox zones in the EBPR cycle ^{10,102}. However, it is still unclear which

conditions trigger different modes of TCA operation ^{70,103–105}. Therefore, we investigated all genes related to the proposed anaerobic TCA types of operation, namely: the full, partial (bypass through the glyoxylate shunt) and split modes (right, or the oxidative arm of the TCA cycle running forward and left, or the reductive arm operating backward through the succinate–propionate pathway). Our results show that, although these two genomes encode for all genes related to a full anaerobic TCA operation, they have unique expression patterns, which seem to be regulated upon different environmental and physiological cues. For instance, (1) the upregulation of the full glyoxylate cycle only in *Accumulibacter* Clade IA-UW3, (2) clade-specific regulation of KFOR vs 2-oxoglutarate dehydrogenase, and (3) contrasting regulation of the two citrate synthase genes, explain inconsistencies amongst early investigations. Therefore, the existence of multiple *Accumulibacter* clades in those studies could provide insight on the debate concerning this intriguing aspect of *Accumulibacter*'s physiology.

Nitrogen metabolism is also inferred to drive ecological differences between *Accumulibacter* clades and there is an ongoing debate to determine to which extent members of the *Accumulibacter* lineage can transform nitrate to nitrogen gas ^{12,25,106,107}. Recent genomic investigations revealed full denitrification capabilities of one member of the *Accumulibacter* lineage acclimated to micro-aerobic conditions (*Accumulibacter* Clade IC-UW_{LDO} ⁸¹), and here we present additional evidence for this phenotype in another *Accumulibacter* clade. Our results reveal that the anoxic respiration machinery is activated during the anaerobic phase, implying that *Accumulibacter* IA-UW3 is fully prepared to use electron acceptors different than

oxygen likely due to the transcriptional factors that control these pathways, which has been previously reported ⁸¹.

Importance of substrate concentration during polymer storage

We also paid close attention to all genes related to acetate conversion and its further transformation into PHB. Many authors have overlooked at the specific conditions which activate the low- versus high-affinity acetate uptake routes, which might have repercussions in further molecule signaling and regulation via Acetyl-P production ¹⁰⁸. Acetyl-P is known to act as a global signal in *E. coli*, and it is considered as an ideal global molecule due to its relatively small size, both low energy-cost and half-life and capability of effecting the coordinated regulation of diverse cellular processes ¹⁰⁹. It has been also shown that post-translational control by acetyl phosphate regulates PHB synthase ¹¹⁰, the last step in PHB synthesis in *Accumulibacter*. Thus, we propose a simple mechanism (Figure 5) that can help readers understand the consequences of acetate availability and its concentration with the goal of better process design principles.

Concluding remarks

Transcriptomics experiments can reveal large-scale differences in the immediate response of closely-related microorganisms to equal environmental conditions. However, the impact of such variations in gene expression on the phenotypes of *Accumulibacter* is not yet understood. A bottom-up analysis of changes in fine-

tuned cellular processes and coordination among transcriptional programs in *Accumulibacter* is of great importance to help to solve this challenge.

Acknowledgements

This work was partially supported by funding from the National Science Foundation (MCB-1518130). Additional funding from the Chilean National Commission for Scientific and Technological Research (CONICYT) as a fellowship to Francisco Moya is also acknowledged. Michael Eng Hoe Khor was supported by the 2017 UW-Madison Hilldale Scholarship. We thank the Joint Genome Institute and University of Wisconsin Biotechnology Center for DNA sequencing support, and members of the McMahon Lab for early revisions of the manuscript. The authors also thank Matthew Kizaric, for operation of the bioreactor during the course of the study and Matthew Scarborough for help with RNA processing.

Figures Descriptions

Figure 1 | Accumulibacter Pan-transcriptome in response to classic feast and famine conditions.

Double-layered Venn diagram depicting orthology and differential gene expression (DE) for both Accumulibacter clades enriched under this study. In red, Accumulibacter Clade IA-UW3 features. In blue, Accumulibacter Clade IIC-UW6 features. "Core" genes represent those that share orthology, based on reciprocal best BLAST hits. All the non-core genes are described as "Flexible" genes. See results for more details.

Figure 2 | Functional differences of gene expression at the clade level.

Accumulibacter-core genes and their expression patterns based on functional KEGG classification. We selected orthologous genes that had functional categories associated to them and compared their gene expression profile (i.e. whether those were differentially expressed) and average correlation score across each functional group (Pearson product correlation at $p > 0.1$).

Figure 3 | Gene expression dynamics across the EBPR cycle.

Selection of central-carbon metabolism related genes and their expression pattern over time. Gene expression was normalized based on entropy patterns and depicted as downregulated (blue), constitutively expressed (light and dark gray), and upregulated (green). Genes present in a single clade (flexible) are shaded in white.

Figure 4 | Presence and expression of denitrification genes across Accumulibacter Clades IA-UW3 and IIC-UW6.

Colored arrows denote the phase at which differential gene expression was observed for all enzyme subunits in the cycle.

Figure 5 | Proposed mode of operation for anaerobic acetate uptake and conversion to PHB in Accumulibacter.

Genes colored in blue represent upregulation, and those in red were downregulated during anaerobic acetate addition. Dashed arrows denote activation (>) or repression (|) of an specific gene, upon metabolites' concentration.

Figures

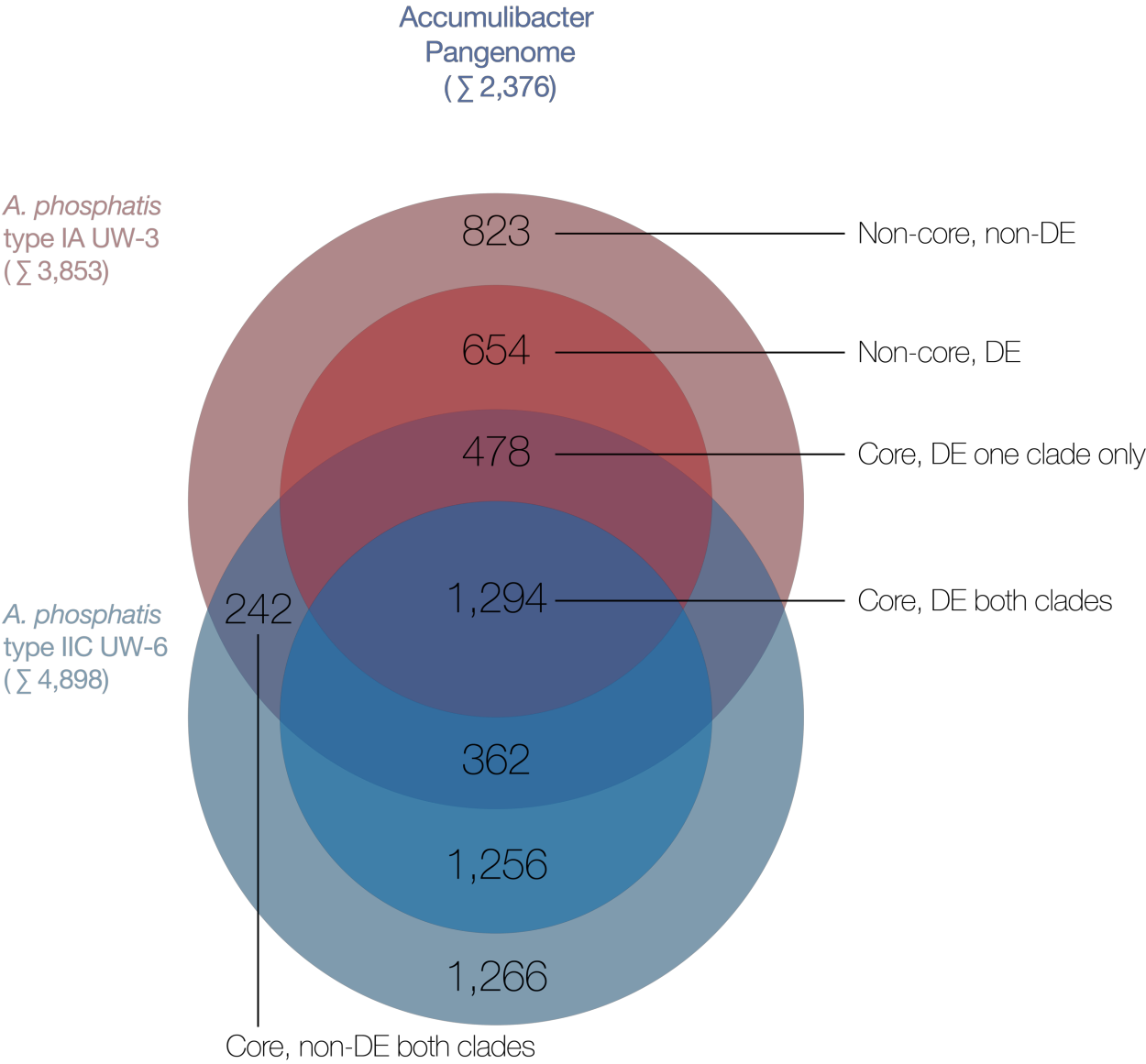


Figure 1

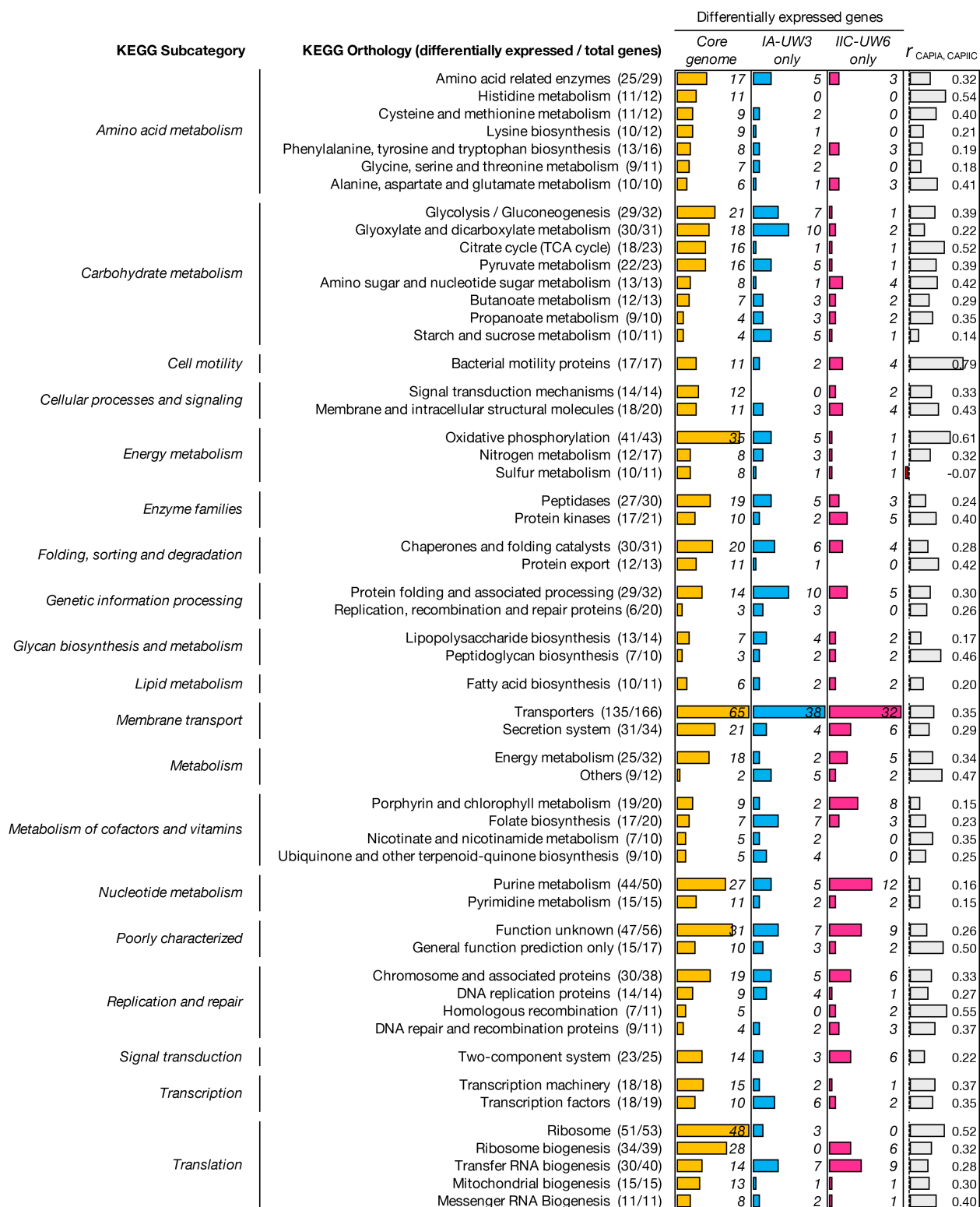


Figure 2

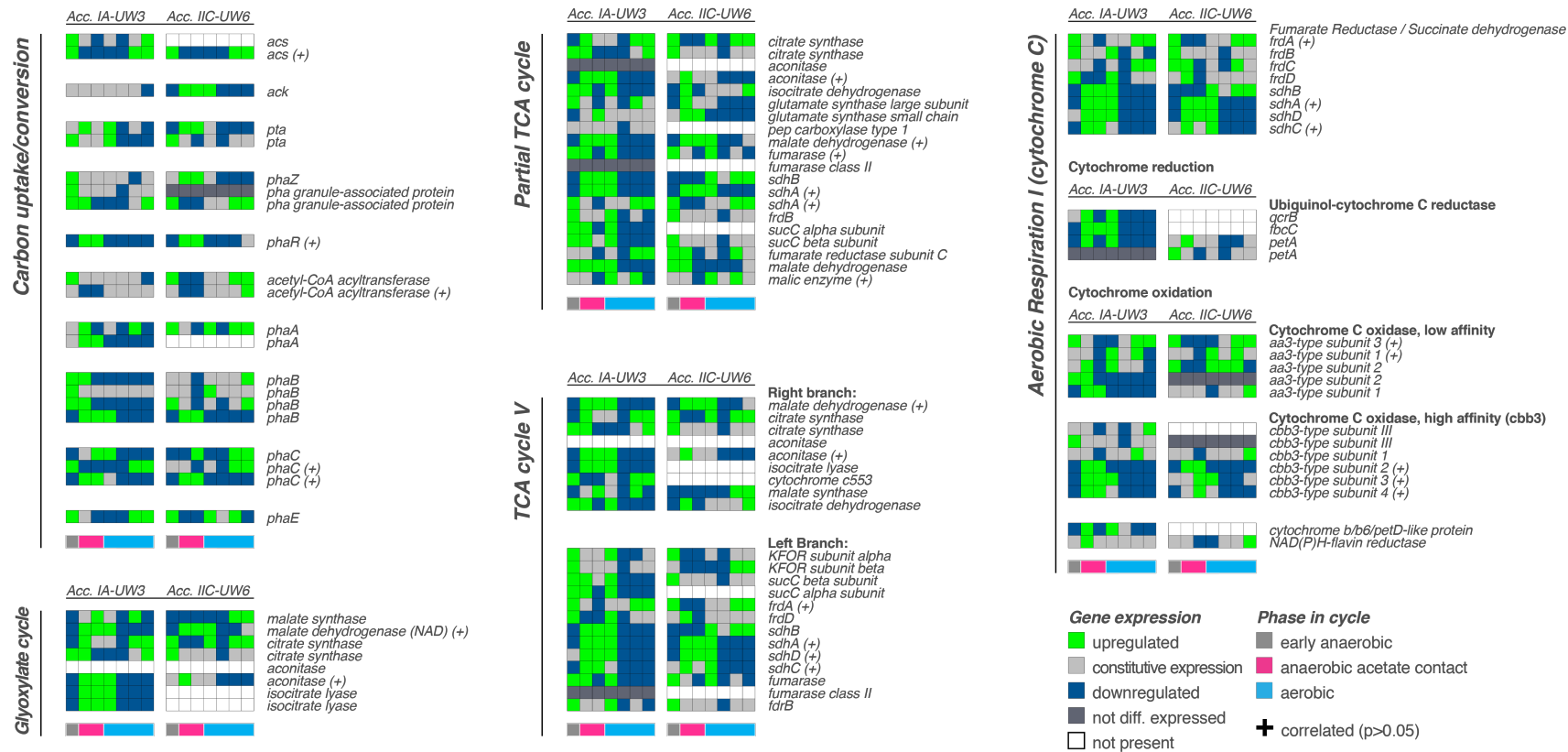


Figure 3

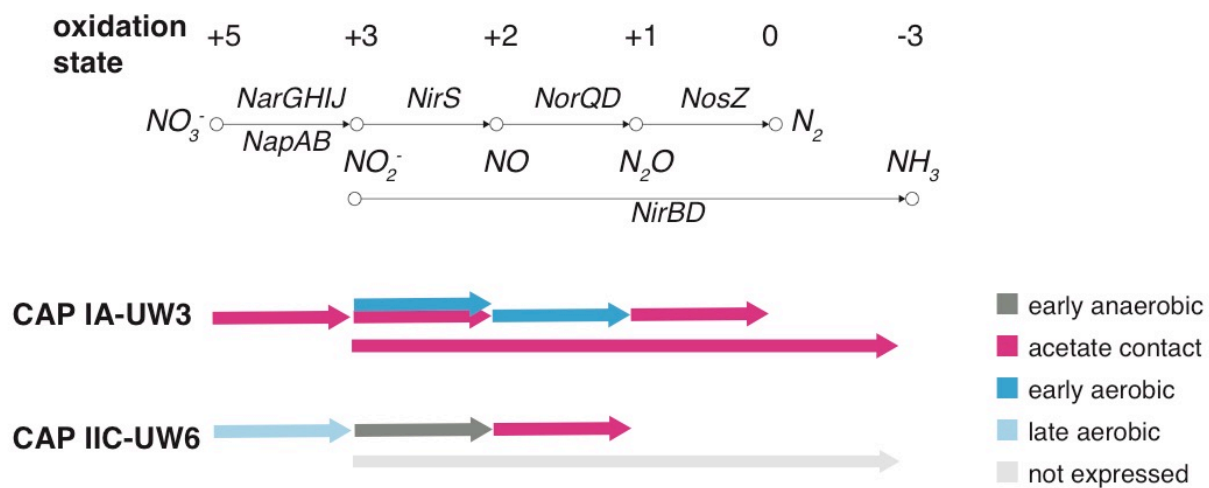


Figure 4

Figure 5

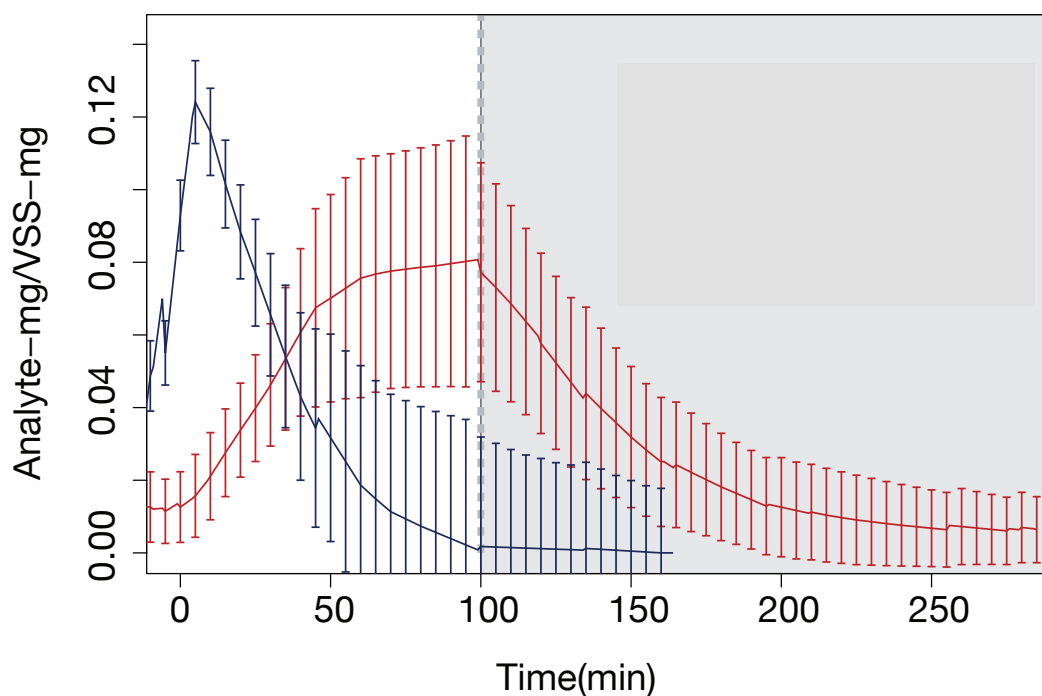
Tables

Table 1 | Metagenome-assembled genomes (MAGs) of *Accumulibacter* obtained from lab-scale reactors in 2016.

Bin Id	Completeness (%)	Redundancy (%)	Genome size (bp)	# scaffolds	# contigs	N50 (scaffolds)	N50 (contigs)
CAP IA-UW4	84.98	4.25	4,291,275	150	441	2,834,893	27,333
CAP IIA-UW5	98.99	5.24	4,882,958	1	68	4,882,958	124,608
CAP IIC-UW6	98.57	2.46	5,179,050	1	1	5,179,050	5,178,808
CAP IIF-UW7	93.98	0.66	4,901,359	377	1,067	31,147	9,538

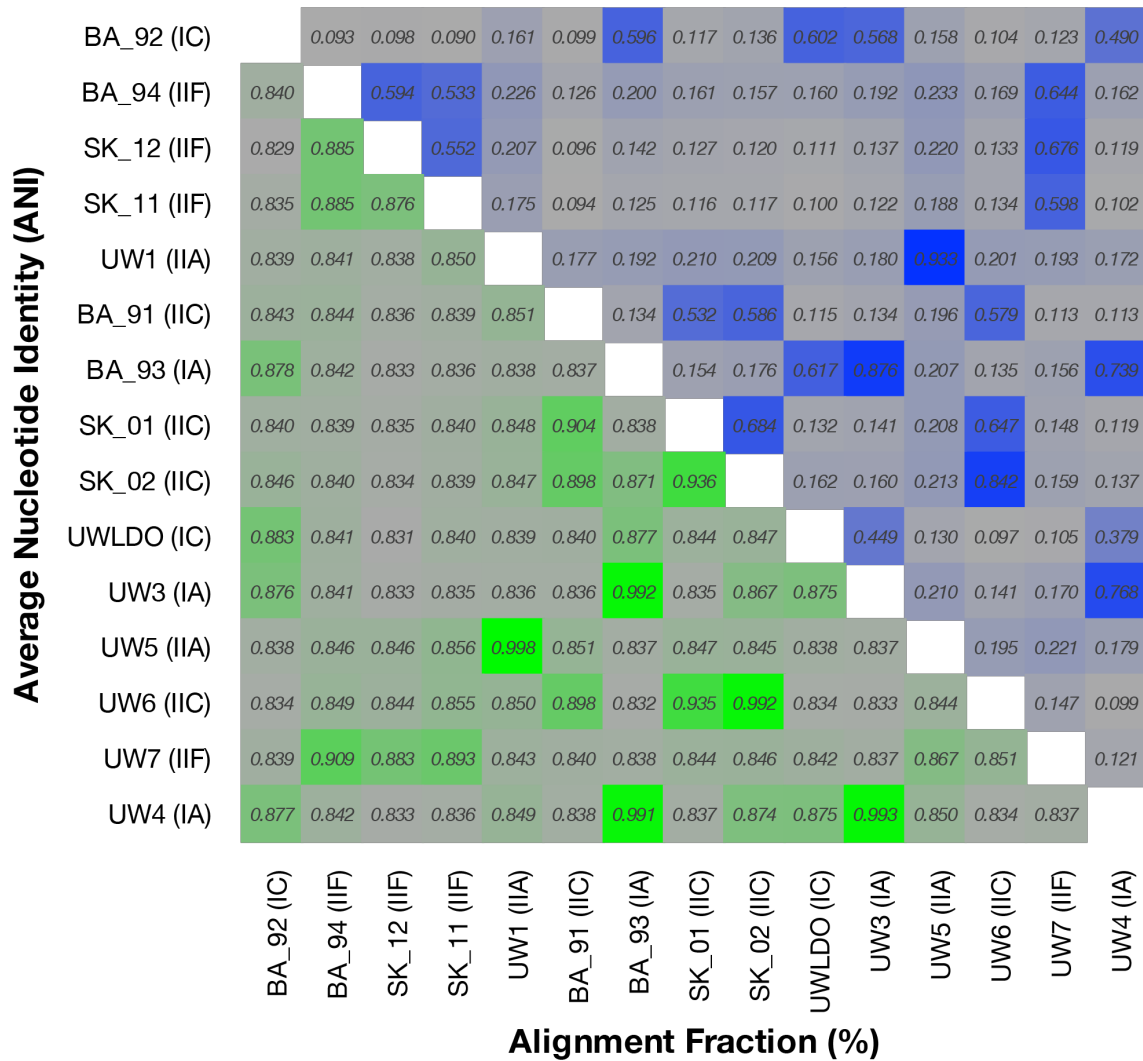
Supplementary Materials

Supplementary Figures



Supplementary Figure S1 | Time-series EBPR data.

Averaged data of P (red) and acetate (blue) profiles of *Accumulibacter*-enriched reactors from its seeding in 2015 to the end of 2016, normalized by VSS, with error bars. White and gray shading represents anaerobic and aerobic phases, respectively.



Supplementary Figure S2 | ANI vs Percent Alingment of several Accumulibacter genomes.

Heatmap showing the average nucleotide identity (lower diagonal) and alignment fraction (upper diagonal) for a collection of Accumulibacter genomes with known phylogeny and the ones obtained in this study (UW4-UW7).

Supplementary Tables

Supplementary Table S1 | Metatranscriptomics reads analysis.

Sample time	Phase in cycle	FASTQ	FASTA - post QC				
		Reads	Reads	rRNA	mRNA	FASTQ norm	FASTA norm
10:45:00 AM	Early anaerobic	25,667,164	18,193,544	8,039,709	10,153,835	1.480	1.551
11:16:00 AM	Ana. acetate contact	32,437,715	23,015,194	10,430,238	12,584,956	1.171	1.252
11:55:00 AM	Ana. acetate contact	35,549,288	24,565,614	10,939,385	13,626,229	1.069	1.156
12:40:00 PM	Aerobic	37,988,595	25,499,475	9,748,766	15,750,709	1.000	1.000
1:15:00 PM	Aerobic	31,364,122	21,774,407	10,500,524	11,273,883	1.211	1.397
1:55:00 PM	Aerobic	28,344,123	20,259,579	8,372,740	11,886,839	1.340	1.325
2:55:00 PM	Aerobic	28,016,344	18,383,083	9,853,738	8,529,345	1.356	1.847

Supplementary Table S2 | Accumulibacter Pan-transcriptome information.

	Total genes	Flexible, non-DE	Flexible, DE	Core, clade-specific DE	Core, other clade DE	Core, DE	Core, non-DE
Accumulibacter IA-UW3	3853	21.4%	17.0%	12.4%	9.4%	33.6%	6.3%
Accumulibacter IIC-UW6	4898	25.8%	25.6%	7.4%	9.8%	26.4%	4.9%

Supplementary Table S3 | mRNA reads mapped to all four Accumulibacter clades present in the chemostat.

Phase in cycle / Accumulibacter clade	Early anaerobic	Ana. acetate contact t ₁	Ana. acetate contact t ₂	Aerobic t ₁	Aerobic t ₂	Aerobic t ₃	Aerobic t ₄	Σ mRNA reads	% mRNA
IA-UW3	955,920	1,455,359	1,496,778	2,057,294	1,210,289	1,014,083	633,204	8,822,927	8.60%
IIA-UW5	63,551	99,767	96,019	103,715	68,022	72,487	54,058	557,619	0.58%
IIC-UW6	3,254,129	3,670,084	4,147,908	4,595,586	2,976,412	4,329,626	2,272,117	25,245,862	27.57%
IIF-UW7	22,762	69,999	91,649	99,793	43,705	36,854	20,951	385,713	0.44%
								Σ Accumulibacter reads	37.20%

Supplementary Data Table S4 (next 6 pages) | Gene expression dynamics analyzed under this study.

Genes and their expression pattern over time. Gene expression was normalized based on entropy patterns and depicted as downregulated (-1), constitutively expressed (0), and upregulated (1). N/A genes did not present orthologs in one of the two genomes analyzed.

Genes/Pathways

Aerobic Respiration I (cytochrome C)

Acc type IA UW3 (Ga0131788)

Control_1045
Control_1116
Control_1155
Control_1240
Control_1315
Control_1355
Control_1455

Succinate to fumarate

Fumarate Reductase / Succinate dehydrogenase

Subunits

Locus tag

Ga0131788_111667	<i>succinate dehydrogenase subunit A</i>	1	0	-1	0	0	1	1
Ga0131788_111668	<i>fumarate reductase iron-sulfur subun</i>	1	0	0	1	-1	0	-1
Ga0131788_111669	<i>fumarate reductase subunit C</i>	0	0	-1	0	-1	1	1
Ga0131788_111670	<i>succinate dehydrogenase subunit D</i>	1	-1	-1	1	-1	0	0
Ga0131788_12494	<i>succinate dehydrogenase subunit B</i>	-1	1	1	1	-1	-1	-1
Ga0131788_12495	<i>succinate dehydrogenase subunit A</i>	-1	1	1	1	-1	-1	-1
Ga0131788_12496	<i>succinate dehydrogenase subunit D</i>	0	1	1	1	-1	-1	-1
Ga0131788_12497	<i>succinate dehydrogenase subunit C</i>	-1	1	1	0	-1	-1	-1

Acc type IIC UW6 (Ga0248264)

Control_1045
Control_1116
Control_1155
Control_1240
Control_1315
Control_1355
Control_1455

Locus tag

Ga0248264_111802	<i>succinate dehydrogenase subunit A</i>	1	-1	-1	0	0	1	1
Ga0248264_111803	<i>succinate dehydrogenase subunit B</i>	1	0	0	0	-1	0	0
Ga0248264_111987	<i>fumarate reductase subunit C</i>	1	1	-1	0	-1	1	0
Ga0248264_111986	<i>succinate dehydrogenase subunit D</i>	0	1	-1	0	0	0	0
Ga0248264_112031	<i>succinate dehydrogenase subunit B</i>	-1	-1	-1	1	0	1	1
Ga0248264_112032	<i>succinate dehydrogenase subunit A</i>	-1	1	1	1	-1	-1	-1
Ga0248264_112033	<i>succinate dehydrogenase subunit D</i>	-1	1	1	1	-1	-1	-1
Ga0248264_112034	<i>succinate dehydrogenase subunit C</i>	0	1	0	1	-1	-1	-1

NADH to NAD

NADH dehydrogenase

Subunits

Ga0131788_11347	<i>NADH dehydrogenase subunit N</i>	-1	0	0	1	0	0	0
Ga0131788_11348	<i>NADH dehydrogenase subunit M</i>	-1	0	1	0	0	0	0
Ga0131788_11349	<i>NADH dehydrogenase subunit L</i>	-1	0	1	0	0	0	-1
Ga0131788_11350	<i>NADH dehydrogenase subunit K</i>	-1	1	1	-1	-1	0	0
Ga0131788_11351	<i>NADH dehydrogenase subunit J</i>	-1	0	1	1	1	0	0
Ga0131788_11352	<i>NADH dehydrogenase subunit I</i>	0	-1	0	1	1	0	0
Ga0131788_11353	<i>NADH dehydrogenase subunit H</i>	-1	1	1	0	0	0	-1
Ga0131788_11354	<i>NADH dehydrogenase subunit G</i>	-1	1	1	0	-1	-1	-1
Ga0131788_11355	<i>NADH dehydrogenase subunit F</i>	-1	1	1	1	-1	-1	-1
Ga0131788_11356	<i>NADH dehydrogenase subunit E</i>	-1	1	1	1	0	-1	-1
Ga0131788_11357	<i>NADH dehydrogenase subunit D</i>	-1	1	1	1	-1	-1	-1
Ga0131788_11358	<i>NADH dehydrogenase subunit C</i>	-1	1	0	1	-1	-1	-1
Ga0131788_11359	<i>NADH dehydrogenase subunit B</i>	-1	1	1	1	-1	-1	-1
Ga0131788_11360	<i>NADH dehydrogenase subunit A</i>	-1	1	1	1	-1	-1	-1
Ga0131788_11501	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0
Ga0131788_11504	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0
Ga0131788_11505	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0

Ga0248264_11885	<i>NADH dehydrogenase subunit N</i>	-1	0	0	0	0	0	0
Ga0248264_11884	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0
Ga0248264_11883	<i>NADH dehydrogenase subunit L</i>	0	0	0	0	0	0	0
Ga0248264_11882	<i>NADH dehydrogenase subunit K</i>	-1	1	0	0	0	0	0
Ga0248264_11881	<i>NADH dehydrogenase subunit J</i>	0	0	-1	0	0	0	0
Ga0248264_11880	<i>NADH dehydrogenase subunit I</i>	-1	1	0	0	0	0	0
Ga0248264_11879	<i>NADH dehydrogenase subunit H</i>	-1	1	1	0	-1	0	-1
Ga0248264_11878	<i>NADH dehydrogenase subunit G</i>	-1	1	1	1	-1	-1	-1
Ga0248264_11877	<i>NADH dehydrogenase subunit F</i>	-1	1	1	1	-1	-1	-1
Ga0248264_11876	<i>NADH dehydrogenase subunit E</i>	-1	0	0	1	0	0	-1
Ga0248264_11875	<i>NADH dehydrogenase subunit D</i>	-1	1	1	1	-1	-1	-1
Ga0248264_11874	<i>NADH dehydrogenase subunit C</i>	0	1	0	1	-1	-1	-1
Ga0248264_11873	<i>NADH dehydrogenase subunit B</i>	1	0	-1	1	0	1	-1
Ga0248264_11872	<i>NADH dehydrogenase subunit A</i>	0	1	0	1	-1	-1	-1
Ga0248264_111638	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0
Ga0248264_111635	<i>NADH dehydrogenase subunit M</i>	0	0	0	0	0	0	0
Ga0248264_111634	<i>NADH dehydrogenase subunit M</i>	1	0	0	0	-1	0	1

Cytochrome reduction

Ubiquinol-cytochrome C reductase

Subunits

Ga0131788_11946	<i>ubiquinol-cytochrome c reductase cy</i>	0	1	-1	1	-1	-1	-1
Ga0131788_11948	<i>ubiquinol-cytochrome c reductase cy</i>	-1	1	1	1	-1	-1	-1
Ga0131788_11945	<i>ubiquinol-cytochrome c reductase ir</i>	-1	1	-1	1	-1	-1	-1

Ga0248264_114009	<i>ubiquinol-cytochrome c reductase cy</i>	0	0	0	0	0	0	0
Ga0248264_114008	<i>ubiquinol-cytochrome c reductase cy</i>	0	0	0	0	0	0	0
Ga0248264_11574	<i>ubiquinol-cytochrome c reductase ir</i>	0	1	0	0	-1	-1	0
Ga0248264_113744	<i>ubiquinol-cytochrome c reductase ir</i>	1	0	-1	0	-1	0	0

Cytochrome oxidation

Cytochrome C oxidase_low affinity

Subunits

Ga0131788_11742	<i>cytochrome c oxidase subunit 3</i>	1	0	-1	-1	0	1	1
Ga0131788_11746	<i>cytochrome c oxidase subunit 1</i>	0	0	-1	1	0	1	-1
Ga0131788_11747	<i>cytochrome c oxidase subunit 2</i>	0	1	-1	1	0	0	-1
Ga0131788_12643	<i>cytochrome c oxidase subunit 2</i>	1	1	-1	-1	-1	-1	-1
Ga0131788_12644	<i>cytochrome c oxidase subunit 1</i>	-1	1	1	-1	-1	-1	-1

Ga0248264_11569	<i>cytochrome c oxidase subunit 3</i>	1	-1	-1	-1	0	1	1
Ga0248264_11573	<i>cytochrome c oxidase subunit 1</i>	0	0	-1	1	0	1	0
Ga0248264_11574	<i>cytochrome c oxidase subunit 2</i>	1	-1	-1	1	1	1	-1
N/A	N/A							
Ga0248264_112711	<i>cytochrome c oxidase subunit 1</i>	0	0	0	-1	0	0	1

Cytochrome C oxidase_high affinity (cbh3)

Subunits

Ga0131788_11786	<i>cbh3-type cytochrome c oxidase sub</i>	0	0	-1	0	-1	0	1
Ga0131788_111180	<i>cbh3-type cytochrome c oxidase sub</i>	1	0	0	0	-1	0	0
Ga0131788_112380	<i>cytochrome c oxidase cbh3-type sub</i>	0	0	-1	0	0	1	0
Ga0131788_112381	<i>cytochrome c oxidase cbh3-type sub</i>	-1	1	1	-1	-1	-1	-1
Ga0131788_112383	<i>cytochrome c oxidase cbh3-type sub</i>	-1	1	1	1	-1	-1	-1
Ga0131788_112382	<i>cytochrome c oxidase cbh3-type sub</i>	-1	1	1	0	-1	-1	-1
Ga0131788_11947	<i>cytochrome b/b6/petD-like protein</i>	-1	1	-1	1	0	-1	-1
Ga0131788_12650	<i>NAD(P)H-flavin reductase</i>	0	1	0	0	0	0	0

Ga0248264_113108	<i>mono/diheme cytochrome c family p</i>	0	0	0	0	0	0	0
N/A	N/A							
Ga0248264_111826	<i>cytochrome c oxidase cbh3-type sub</i>	0	0	-1	0	0	0	1
Ga0248264_111827	<i>cytochrome c oxidase cbh3-type sub</i>	-1	1	1	-1	-1	-1	-1
Ga0248264_111829	<i>cytochrome c oxidase cbh3-type sub</i>	0	0	1	1	-1	-1	0
Ga0248264_111828	<i>cytochrome c oxidase cbh3-type sub</i>	-1	0	1	0	-1	-1	-1
Ga0248264_114009	<i>ubiquinol-cytochrome c reductase cy</i>	0	0	0	0	0	0	0
Ga0248264_112717	<i>quinol-cytochrome oxidoreductase c</i>	0	0	0	-1	-1	0	1

Novel cytochrome

Genes/Pathways		Acc type IA UW3 (Ga0131788)		Acc type IIC UW6 (Ga0248264)																
Anaerobic Respiration (Glycerol-3-phosphate to fumarate electron transfer)																				
Fumarate to succinate																				
<u>Fumarate reductase / succinate dehydrogenase</u>		Locus tag		Locus tag																
Subunits	Ga0131788_111667	succinate dehydrogenase subunit A	<table><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	1	0	-1	0	0	1	1	Ga0248264_111802	succinate dehydrogenase subunit A	<table><tr><td>1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	1	-1	-1	0	0	1	1
1	0	-1	0	0	1	1														
1	-1	-1	0	0	1	1														
	Ga0131788_111668	fumarate reductase iron-sulfur subunit	<table><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td></tr></table>	1	0	0	1	-1	0	-1	Ga0248264_111803	succinate dehydrogenase subunit B	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	1	0	0	0	-1	0	0
1	0	0	1	-1	0	-1														
1	0	0	0	-1	0	0														
	Ga0131788_111669	fumarate reductase subunit C	<table><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>1</td></tr></table>	0	0	-1	0	-1	1	1	Ga0248264_111987	fumarate reductase subunit C	<table><tr><td>1</td><td>1</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>0</td></tr></table>	1	1	-1	0	-1	1	0
0	0	-1	0	-1	1	1														
1	1	-1	0	-1	1	0														
	Ga0131788_12494	succinate dehydrogenase subunit B	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_112031	succinate dehydrogenase subunit B	<table><tr><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	-1	-1	-1	1	0	1	1
-1	1	1	1	-1	-1	-1														
-1	-1	-1	1	0	1	1														
	Ga0131788_12495	succinate dehydrogenase subunit A	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_112032	succinate dehydrogenase subunit A	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
Anaerobic Respiration (Hydrogen to fumarate electron transfer)																				
Menaquinol to menaquinone																				
<u>Hydrogenase</u>																				
Subunits:	Ga0131788_13188	nickel-dependent hydrogenase	<table><tr><td>0</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	1	-1	-1	-1	-1	N/A	N/A	<table><tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>							
0	1	1	-1	-1	-1	-1														
	Ga0131788_112212	hydrogenase large subunit	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	Ga0248264_112100	hydrogenase large subunit	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0
0	0	0	0	0	0	0														
0	0	0	0	0	0	0														
	Ga0131788_112215	hydrogenase small subunit	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	Ga0248264_112103	hydrogenase small subunit	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0
0	0	0	0	0	0	0														
0	0	0	0	0	0	0														
	Ga0131788_111369	NAD(P)-dependent nickel-iron dehydrogenase	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1	Ga0248264_113465	nickel-dependent hydrogenase	<table><tr><td>0</td><td>1</td><td>-1</td><td>1</td><td>0</td><td>-1</td><td>-1</td></tr></table>	0	1	-1	1	0	-1	-1
-1	1	1	-1	-1	-1	-1														
0	1	-1	1	0	-1	-1														
	N/A	N/A		Ga0248264_113521	nickel-dependent hydrogenase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	0	0							
1	0	0	0	0	0	0														
Fumarate to succinate																				
<u>Fumarate reductase / succinate dehydrogenase</u>																				
Subunits:	Ga0131788_111668	fumarate reductase iron-sulfur subunit	<table><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td></tr></table>	1	0	0	1	-1	0	-1	Ga0248264_111803	succinate dehydrogenase subunit B	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	1	0	0	0	-1	0	0
1	0	0	1	-1	0	-1														
1	0	0	0	-1	0	0														
	Ga0131788_111669	fumarate reductase subunit C	<table><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>1</td></tr></table>	0	0	-1	0	-1	1	1	Ga0248264_111987	fumarate reductase subunit C	<table><tr><td>-1</td><td>1</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>0</td></tr></table>	-1	1	-1	0	-1	1	0
0	0	-1	0	-1	1	1														
-1	1	-1	0	-1	1	0														
	Ga0131788_12495	succinate dehydrogenase subunit A	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_112032	succinate dehydrogenase subunit A	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
	Ga0131788_111667	succinate dehydrogenase subunit A	<table><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	1	0	-1	0	0	1	1	Ga0248264_111802	succinate dehydrogenase subunit A	<table><tr><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	-1	-1	-1	0	0	1	1
1	0	-1	0	0	1	1														
-1	-1	-1	0	0	1	1														
	Ga0131788_12494	succinate dehydrogenase subunit B	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_112031	succinate dehydrogenase subunit B	<table><tr><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	-1	-1	-1	1	0	1	1
-1	1	1	1	-1	-1	-1														
-1	-1	-1	1	0	1	1														
2-oxoglutarate decarboxylation to succinyl-CoA (NADH-dependent)																				
Hydrogen production (NADH dependent)																				
<u>Hydrogen production (NADH dependent)</u>																				
Subunits	Ga0131788_111367	[NiFe] hydrogenase diaphorase moiety	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1	Ga0248264_113295	[NiFe] hydrogenase diaphorase moiety	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0
-1	1	1	-1	-1	-1	-1														
0	0	0	0	0	0	0														
	Ga0131788_111368	NAD-reducing hydrogenase small subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1	N/A	N/A								
-1	1	1	-1	-1	-1	-1														
	Ga0131788_111369	NAD(P)-dependent nickel-iron dehydrogenase	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1	Ga0248264_113465	nickel-dependent hydrogenase	<table><tr><td>0</td><td>1</td><td>-1</td><td>1</td><td>0</td><td>-1</td><td>-1</td></tr></table>	0	1	-1	1	0	-1	-1
-1	1	1	-1	-1	-1	-1														
0	1	-1	1	0	-1	-1														
	Ga0131788_111366	NAD(P)-dependent nickel-iron dehydrogenase	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	1	0	-1	-1	-1	Ga0248264_113294	NAD(P)-dependent nickel-iron dehydrogenase	<table><tr><td>0</td><td>0</td><td>0</td><td></td><td></td><td></td><td></td></tr></table>	0	0	0				
0	1	1	0	-1	-1	-1														
0	0	0																		
	Ga0131788_111370	hydrogenase maturation protease	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_113464	hydrogenase maturation protease	<table><tr><td>-1</td><td>1</td><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td></tr></table>	-1	1	-1	1	1	0	-1
-1	1	1	1	-1	-1	-1														
-1	1	-1	1	1	0	-1														
	Ga0131788_13182	hydrogenase expression/formation protein	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	Ga0248264_113496	hydrogenase expression/formation protein	<table><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	-1	0	0	0	0
0	0	0	0	0	0	0														
0	0	-1	0	0	0	0														
	Ga0131788_13184	hydrogenase expression/formation protein	<table><tr><td>0</td><td>0</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>0</td></tr></table>	0	0	1	1	-1	-1	0	Ga0248264_113517	hydrogenase expression/formation protein	<table><tr><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	-1	-1	-1	0	0	1	1
0	0	1	1	-1	-1	0														
-1	-1	-1	0	0	1	1														
	Ga0131788_13185	hydrogenase maturation protein HypB	<table><tr><td>0</td><td>0</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	0	1	0	-1	-1	-1	Ga0248264_113518	hydrogenase maturation protein HypB	<table><tr><td>-1</td><td>0</td><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td></tr></table>	-1	0	0	1	-1	0	-1
0	0	1	0	-1	-1	-1														
-1	0	0	1	-1	0	-1														
ATP synthesis																				
<u>ATP Synthase</u>																				
Subunits	Ga0131788_11850	ATP synthase F0 subcomplex C subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11364	ATP synthase F0 subcomplex C subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
	Ga0131788_11847	ATP synthase F1 subcomplex alpha subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11367	ATP synthase F1 subcomplex alpha subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
	Ga0131788_11845	ATP synthase F1 subcomplex beta subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11369	ATP synthase F1 subcomplex beta subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
	Ga0131788_11848	ATP synthase F1 subcomplex delta subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11366	ATP synthase F1 subcomplex delta subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	0	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	0	-1	-1	-1														
	Ga0131788_11844	ATP synthase F1 subcomplex epsilon subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	0	-1	-1	-1	Ga0248264_11370	ATP synthase F1 subcomplex epsilon subunit	<table><tr><td>-1</td><td>0</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	0	1	0	-1	-1	-1
-1	1	1	0	-1	-1	-1														
-1	0	1	0	-1	-1	-1														
	Ga0131788_11846	ATP synthase F1 subcomplex gamma subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11368	ATP synthase F1 subcomplex gamma subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														
	Ga0131788_11852	ATP synthase protein I	<table><tr><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td></tr></table>	-1	-1	-1	1	1	-1	-1	Ga0248264_11362	ATP synthase protein I	<table><tr><td>-1</td><td>1</td><td>-1</td><td>1</td><td>0</td><td>0</td><td>-1</td></tr></table>	-1	1	-1	1	0	0	-1
-1	-1	-1	1	1	-1	-1														
-1	1	-1	1	0	0	-1														
	Ga0131788_112438	F-type H ⁺ -transporting ATPase subunit	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr></table>	1	0	0	0	0	0	-1	N/A	N/A								
1	0	0	0	0	0	-1														
	Ga0131788_111607	flagellum-specific ATP synthase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr></table>	1	0	0	0	0	0	-1	Ga0248264_11849	flagellum-specific ATP synthase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0
1	0	0	0	0	0	-1														
0	0	0	0	0	0	0														
	Ga0131788_111841	H ⁺ -transporting ATPase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>-1</td></tr></table>	0	0	0	0	-1	0	-1	N/A	N/A								
0	0	0	0	-1	0	-1														
	Ga0131788_11851	F-type H ⁺ -transporting ATPase subunit	<table><tr><td>-1</td><td>1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	-1	1	-1	-1	-1	Ga0248264_11363	F-type H ⁺ -transporting ATPase subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	0	-1	-1	-1
-1	1	-1	1	-1	-1	-1														
-1	1	1	0	-1	-1	-1														
	Ga0131788_11849	F-type H ⁺ -transporting ATPase subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11365	F-type H ⁺ -transporting ATPase subunit	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1
-1	1	1	1	-1	-1	-1														
-1	1	1	1	-1	-1	-1														

Genes/Pathways		Acc type IA UW3 (Ga0131788)			Acc type IIC UW6 (Ga0248264)																														
Acetate conversion / transport																																			
Acetyl-CoA synthetase		Locus tag			Locus tag																														
Subunits	Ga0131788_111520	acetyl-CoA synthetase	<table><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr></table>	1	0	-1	0	-1	0	1	1	-1	-1	-1	-1	1	1	Ga0248264_113037	acetyl-CoA synthetase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr></table>	0	0	0	0	0	0	0	1	-1	-1	-1	-1	1	1	
1	0	-1	0	-1	0	1																													
1	-1	-1	-1	-1	1	1																													
0	0	0	0	0	0	0																													
1	-1	-1	-1	-1	1	1																													
	Ga0131788_11333	acetyl-coenzyme A synthetase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr></table>	1	0	0	0	0	0	-1	Ga0248264_11705	acetyl-coenzyme A synthetase	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1															
1	0	0	0	0	0	-1																													
-1	1	1	1	-1	-1	-1																													
Acetate kinase		Ga0131788_12393	acetate kinase	<table><tr><td>0</td><td>1</td><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	0	1	-1	0	-1	1	0	0	1	-1	-1	-1	Ga0248264_112852	acetate kinase	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1							
0	1	0	1	-1	0	-1																													
1	0	0	1	-1	-1	-1																													
-1	1	1	1	-1	-1	-1																													
Phosphate acetyltransferase		Ga0131788_12398	phosphate acetyltransferase	<table><tr><td>0</td><td>1</td><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	0	1	-1	0	-1	1	0	0	1	-1	-1	-1	Ga0248264_112851	phosphate butyryltransferase	<table><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	-1	1	1	0	-1	-1	-1	1	0	-1	0	-1	0	0
0	1	0	1	-1	0	-1																													
1	0	0	1	-1	-1	-1																													
-1	1	1	0	-1	-1	-1																													
1	0	-1	0	-1	0	0																													
	Ga0131788_111365	phosphate acetyltransferase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	1	0	0	0	-1	0	0	1	0	-1	0	-1	0	0	Ga0248264_113293	phosphate butyryltransferase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	1	0	-1	0	-1	0	0	
1	0	0	0	-1	0	0																													
1	0	-1	0	-1	0	0																													
0	0	0	0	0	0	0																													
1	0	-1	0	-1	0	0																													
Propionate CoA-transferase		Ga0131788_12582	propionate CoA-transferase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	1	0	0	0	-1	0	0	1	0	-1	0	-1	0	0	Ga0248264_113623	propionate CoA-transferase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	1	0	-1	0	-1	0	0
1	0	0	0	-1	0	0																													
1	0	-1	0	-1	0	0																													
0	0	0	0	0	0	0																													
1	0	-1	0	-1	0	0																													
	Ga0131788_11216	propionate CoA-transferase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0	1	0	-1	0	-1	0	0	N/A	N/A	<table><tr><td>1</td><td>0</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr></table>	1	0	1	0	-1	-1	0								
0	0	0	0	0	0	0																													
1	0	-1	0	-1	0	0																													
1	0	1	0	-1	-1	0																													
Phosphate butyryltransferase																																			
HpcH/HpaI aldolase/citrate lyase family protein		Ga0131788_112409	HpcH/HpaI aldolase/citrate lyase fam	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	0	0	0	0	-1	0	0	Ga0248264_111886	HpcH/HpaI aldolase/citrate lyase fam	<table><tr><td>1</td><td>0</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr></table>	1	0	1	0	-1	-1	0														
0	0	0	0	-1	0	0																													
1	0	1	0	-1	-1	0																													
Acyl phosphatase																																			
PHB Synthesis - degradation																																			
	Ga0131788_112218	poly(hydroxyalkanoate) depolymerase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	-1	0	1	0	0	0	-1	0	0	Ga0248264_114047	poly(hydroxyalkanoate) depolymerase	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	1	0	-1	-1	-1	1	-1	-1	0	0	1	1	
1	0	0	0	0	-1	0																													
1	0	0	0	-1	0	0																													
0	1	1	0	-1	-1	-1																													
1	-1	-1	0	0	1	1																													
	Ga0131788_112466	poly(hydroxyalkanoate) granule-assc	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>1</td></tr></table>	1	0	0	0	-1	0	0	1	1	-1	-1	-1	0	1	N/A	N/A	<table><tr><td>1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	1	-1	-1	0	0	1	1								
1	0	0	0	-1	0	0																													
1	1	-1	-1	-1	0	1																													
1	-1	-1	0	0	1	1																													
	Ga0131788_1489	poly(hydroxyalkanoate) granule-assc	<table><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>1</td></tr></table>	1	1	-1	-1	-1	0	1	Ga0248264_111237	poly(hydroxyalkanoate) granule-assc	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td></tr></table>	-1	1	1	-1	-1	-1	0															
1	1	-1	-1	-1	0	1																													
-1	1	1	-1	-1	-1	0																													
PHA synthesis repressor PhaR		Ga0131788_112187	polyhydroxyalkanoate synthesis rep	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1	Ga0248264_11760	polyhydroxyalkanoate synthesis rep	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td></tr></table>	-1	1	1	-1	-1	-1	0														
-1	1	1	-1	-1	-1	-1																													
-1	1	1	-1	-1	-1	0																													
b-oxidation		Ga0131788_112045	acetyl-CoA acyltransferase	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr><tr><td>0</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	0	-1	0	-1	-1	0	0	0	0	Ga0248264_111988	acetyl-CoA acyltransferase	<table><tr><td>1</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr><tr><td>0</td><td>-1</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	1	-1	-1	0	0	1	1	0	-1	-1	0	0	0	1
1	0	0	0	0	0	-1																													
0	-1	-1	0	0	0	0																													
1	-1	-1	0	0	1	1																													
0	-1	-1	0	0	0	1																													
	Ga0131788_12452	acetyl-CoA acyltransferase	<table><tr><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>-1</td></tr><tr><td>0</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	-1	0	-1	1	-1	0	1	1	-1	-1	-1	-1	Ga0248264_111200	acetyl-CoA acetyltransferase family	<table><tr><td>1</td><td>0</td><td>-1</td><td>1</td><td>-1</td><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	-1	1	-1	1	1	0	0	0	0	0	0	0	
0	1	-1	0	-1	1	-1																													
0	1	1	-1	-1	-1	-1																													
1	0	-1	1	-1	1	1																													
0	0	0	0	0	0	0																													
phaA		Ga0131788_112231	acetyl-CoA C-acetyltransferase	<table><tr><td>0</td><td>1</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>-1</td></tr><tr><td>0</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	0	1	-1	0	-1	1	-1	0	1	1	-1	-1	-1	-1	Ga0248264_111320	acetyl-CoA C-acetyltransferase	<table><tr><td>1</td><td>0</td><td>-1</td><td>1</td><td>-1</td><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	-1	1	-1	1	1	0	0	0	0	0	0	0
0	1	-1	0	-1	1	-1																													
0	1	1	-1	-1	-1	-1																													
1	0	-1	1	-1	1	1																													
0	0	0	0	0	0	0																													
	Ga0131788_12249	acetyl-CoA C-acetyltransferase	<table><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	Ga0248264_112315	acetyl-CoA C-acetyltransferase	<table><tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	0	0	0	0	0								
1	1	-1	-1	-1	-1	-1																													
1	1	1	-1	-1	-1	-1																													
0	0	0	0	0	0	0																													
phaB		Ga0131788_112044	3-hydroxyacyl-CoA dehydrogenase	<table><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr></table>	1	1	-1	-1	-1	-1	-1	1	0	0	0	0	0	0	Ga0248264_111208	3-hydroxyacyl-CoA dehydrogenase	<table><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr></table>	0	0	-1	0	0	0	1	0	0	-1	1	0	0	0
1	1	-1	-1	-1	-1	-1																													
1	0	0	0	0	0	0																													
0	0	-1	0	0	0	1																													
0	0	-1	1	0	0	0																													
	Ga0131788_112042	3-hydroxyacyl-CoA dehydrogenase/e	<table><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	Ga0248264_111189	3-hydroxyacyl-CoA dehydrogenase/e	<table><tr><td>1</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	0	-1	0	-1	0	1	1	1	1	-1	-1	-1	-1	
1	1	-1	-1	-1	-1	-1																													
1	1	-1	-1	-1	-1	-1																													
1	0	-1	0	-1	0	1																													
1	1	1	-1	-1	-1	-1																													
	Ga0131788_112044	3-hydroxyacyl-CoA dehydrogenase	<table><tr><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	Ga0248264_111212	3-hydroxyacyl-CoA dehydrogenase	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1								
1	1	-1	-1	-1	-1	-1																													
1	1	1	-1	-1	-1	-1																													
-1	1	1	-1	-1	-1	-1																													
	Ga0131788_112186	3-oxoacyl-[acyl-carrier-protein] reduc	<table><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	1	-1	-1	-1	Ga0248264_11759	3-oxoacyl-[acyl-carrier-protein] reduc	<table><tr><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	-1	-1	-1	-1															
-1	1	1	1	-1	-1	-1																													
-1	1	1	-1	-1	-1	-1																													
phaC		Ga0131788_112220	hypothetical protein	<table><tr><td>-1</td><td>0</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr></table>	-1	0	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	Ga0248264_111315	poly(3-hydroxyalkanoate) synthetase	<table><tr><td>-1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>-1</td><td>1</td><td>1</td></tr></table>	-1	-1	1	-1	-1	1	1	0	0	-1	0	-1	1	1
-1	0	1	1	-1	-1	-1																													
-1	-1	-1	-1	-1	1	1																													
-1	-1	1	-1	-1	1	1																													
0	0	-1	0	-1	1	1																													
	Ga0131788_112235	polyhydroxyalkanoate synthase	<table><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>-1</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	-1	1	1	0	-1	-1	-1	-1	1	1	0	-1	-1	-1	Ga0248264_111324	polyhydroxyalkanoate synthase	<table><tr><td>1</td><td>-1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	-1	1	-1	-1	-1	-1								
-1	1	1	0	-1	-1	-1																													
-1	1	1	0	-1	-1	-1																													
1	-1	1	-1	-1	-1	-1																													
	Ga0131788_112223	polyhydroxyalkanoate synthase	<table><tr><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr></table>	1	0	-1	-1	-1	1	1	1	0	-1	-1	-1	1	1	Ga0248264_111317	polyhydroxyalkanoate synthase	<table><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	-1	-1	-1	-1	-1	-1								
1	0	-1	-1	-1	1	1																													
1	0	-1	-1	-1	1	1																													
1	-1	-1	-1	-1	-1	-1																													
	Ga0131788_112234	class III poly(R)-hydroxyalkanoic ac	<table><tr><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td></tr></table>	1	0	-1	-1	-1	1	1	Ga0248264_111323	class III poly(R)-hydroxyalkanoic ac	<table><tr><td>1</td><td>-1</td><td>-1</td><td>1</td><td>0</td><td>1</td><td>-1</td></tr></table>	1	-1	-1	1	0	1	-1															
1	0	-1	-1	-1	1	1																													
1	-1	-1	1	0	1	-1																													
Phosphate Transport / Conversion																																			
	Ga0131788_111284	PiT family inorganic phosphate trans	<table><tr><td>1</td><td>0</td><td>-1</td><td>1</td><td>0</td><td>0</td><td>-1</td></tr><tr><td>-1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	0	-1	1	0	0	-1	-1	1	1	1	-1	-1	-1	Ga0248264_11862	PiT family inorganic phosphate trans	<table><tr><td>0</td><td>-1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	0	-1	1	1	-1	-1	-1	0	0	-1	0	0	0	1	
1	0	-1	1	0	0	-1																													
-1	1	1	1	-1	-1	-1																													
0	-1	1	1	-1	-1	-1																													
0	0	-1	0	0	0	1																													
	Ga0131788_112168	PiT family inorganic phosphate trans	<table><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	-1	-1	-1	1	1	1	1	1	1	1	-1	-1	-1	Ga0248264_11865	PiT family inorganic phosphate trans	<table><tr><td>1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr></table>	1	1	1	1	-1	-1	-1	1	1	1	1	-1	-1	-1	
1	-1	-1	-1	1	1	1																													
1	1	1	1	-1	-1	-1																													
1	1	1	1	-1	-1	-1																													
1	1	1	1	-1	-1	-1																													
	Ga0131788_11313	PiT family inorganic phosphate trans	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr></table>	1	1	0	0	0	0	-1	1	1	0	0	0	0	-1	Ga0248264_112435	PiT family inorganic phosphate trans	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	1	0	-1	-1	0	1	-1	-1	-1	0	1	1	
1	1	0	0	0	0	-1																													
1	1	0	0	0	0	-1																													
0	1	1	0	-1	-1	0																													
1	-1	-1	-1	0	1	1																													
	Ga0131788_12325	PiT family inorganic phosphate trans	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-1</td></tr></table>	1	1	0	0	0	0	-1	1	1	0	0	0	0	-1	N/A	N/A	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	1	0	-1	-1	0	1	-1	-1	-1	0	1	1	
1	1	0	0	0	0	-1																													
1	1	0	0	0	0	-1																													
0	1	1	0	-1	-1	0																													
1	-1	-1	-1	0	1	1																													
	Ga0131788_11181	polyphosphate kinase	<table><tr><td>1</td><td>1</td><td>0</td><td>1</td><td>-1</td><td>-1</td><td>-1</td></tr><tr><td>1</td><td>0</td><td>-1</td><td>-1</td><td>-1</td><td>1</td><td>0</td></tr></table>	1	1	0	1	-1	-1	-1	1	0	-1	-1	-1	1	0	Ga0248264_112421	polyphosphate kinase	<table><tr><td>0</td><td>1</td><td>1</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	1	0	-1	-1	0	1	-1	-1	-1	0	1	1	
1	1	0	1	-1	-1	-1																													
1	0	-1	-1	-1	1	0																													
0	1	1	0	-1	-1	0																													
1	-1	-1	-1	0	1	1																													
	Ga0131788_1171	polyphosphate kinase 2	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	1	1	0	0	-1	-1	0	0	0	-1	-1	0	1	1	Ga0248264_111084	polyphosphate kinase 2	<table><tr><td>1</td><td>-1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr></table>	1	-1	0	0	-1	-1	0	1	-1	0	0	-1	-1	0	
1	1	0	0	-1	-1	0																													
0	0	-1	-1	0	1	1																													
1	-1	0	0	-1	-1	0																													
1	-1	0	0	-1	-1	0																													
	Ga0131788_12103	polyphosphate kinase 2	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	1	1	0	0	-1	-1	0	0	0	-1	-1	0	1	1	Ga0248264_11940	polyphosphate kinase 2	<table><tr><td>1</td><td>-1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>-1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr></table>	1	-1	0	0	-1	-1	0	1	-1	0	0	-1	-1	0	
1	1	0	0	-1	-1	0																													
0	0	-1	-1	0	1	1																													
1	-1	0	0	-1	-1	0																													
1	-1	0	0	-1	-1	0																													
	Ga0131788_12386	polyphosphate kinase 2	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td><td>1</td><td>1</td></tr></table>	1	1	0	0	-1	-1	0	0	0	-1	-1	0	1	1	Ga0248264_112849	polyphosphate kinase 2	<table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>-1</td><td>-1</td><td>0</td></tr></table>	1	1	0	0	-1	-1	0	1	1	0	0	-1	-1	0	
1	1	0	0	-1	-1	0																													
0	0	-1	-1	0	1	1																													
1	1	0	0	-1	-1	0																													
1	1	0	0	-1	-1	0																													

Genes/Pathways	Acc type IA UW3 (Ga0131788)				Acc type IIC UW6 (Ga0248264)
----------------	-----------------------------	--	--	--	------------------------------

TCA cycle: Partial TCA cycle

	Locus tag		Locus tag	
Citrate synthase	Ga0131788_12492	<i>citrate synthase</i>	Ga0248264_112029	<i>citrate synthase</i>
Citrate synthase	Ga0131788_111178	<i>citrate synthase</i>	Ga0248264_113730	<i>citrate synthase</i>
Aconitase	Ga0131788_13107	<i>aconitase</i>	Ga0248264_111436	<i>aconitase</i>
Aconitase	Ga0131788_1222	<i>aconitase</i>	Ga0248264_112069	<i>aconitase</i>
Isocitrate dehydrogenase (NADP)	Ga0131788_13268	<i>isocitrate dehydrogenase (NADP)</i>	Ga0248264_112843	<i>isocitrate dehydrogenase (NADP)</i>
Glutamate synthase (NADH) large subunit	Ga0131788_111333	<i>glutamate synthase (NADH) large subunit</i>	Ga0248264_11515	<i>glutamate synthase (NADH) large subunit</i>
Glutamate synthase (NADPH/NADH) small chain	Ga0131788_111335	<i>glutamate synthase (NADPH/NADH) small chain</i>	Ga0248264_11516	<i>glutamate synthase (NADH) small chain</i>
Phosphoenolpyruvate carboxylase	Ga0131788_1174	<i>phosphoenolpyruvate carboxylase</i>	Ga0248264_111765	<i>phosphoenolpyruvate carboxylase</i>
Malate dehydrogenase (NAD)	Ga0131788_12499	<i>malate dehydrogenase (NAD)</i>	Ga0248264_111563	<i>malate dehydrogenase (NAD)</i>
Fumarase	Ga0131788_11335	<i>homodimeric fumarase (class I)</i>	Ga0248264_11703	<i>homodimeric fumarase (class I)</i>
Fumarase	Ga0131788_112217	<i>fumarase class II</i>	Ga0248264_111932	<i>fumarase hydratase-like protein</i>
Succinate dehydrogenase subunit A	Ga0131788_12494	<i>succinate dehydrogenase subunit B</i>	Ga0248264_112031	<i>succinate dehydrogenase subunit B</i>
Succinate dehydrogenase subunit A	Ga0131788_12495	<i>succinate dehydrogenase subunit A</i>	Ga0248264_112032	<i>succinate dehydrogenase subunit A</i>
Succinate dehydrogenase subunit B	Ga0131788_111667	<i>succinate dehydrogenase subunit A</i>	Ga0248264_111802	<i>succinate dehydrogenase subunit A</i>
Fumarate reductase	Ga0131788_111668	<i>fumarate reductase iron-sulfur subunit</i>	Ga0248264_111803	<i>succinate dehydrogenase subunit B</i>
Succinyl-CoA synthetase (ADP forming) subunit A	Ga0131788_111630	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113615	<i>succinyl-CoA synthetase (ADP-forming)</i>
Succinyl-CoA synthetase (ADP forming) subunit B	Ga0131788_111629	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113616	<i>succinyl-CoA synthetase (ADP-forming)</i>
	Ga0131788_111669	<i>fumarate reductase subunit C</i>	Ga0248264_111987	<i>fumarate reductase subunit C</i>
	Ga0131788_111630	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113638	<i>succinyl-CoA synthetase (ADP-forming)</i>
	Ga0131788_111629	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113639	<i>succinyl-CoA synthetase (ADP-forming)</i>
	Ga0131788_111784	<i>malate dehydrogenase (oxaloacetate)</i>	Ga0248264_113410	<i>malate dehydrogenase (oxaloacetate)</i>
	Ga0131788_111009	<i>allosteric NADP-dependent malic enzyme</i>	Ga0248264_11540	<i>malate dehydrogenase (oxaloacetate)</i>

TCA cycle V (2-oxoglutarate:ferredoxin

Right branch:

Malate dehydrogenase (NAD)	Ga0131788_12499	<i>malate dehydrogenase (NAD)</i>	Ga0248264_111563	<i>malate dehydrogenase (NAD)</i>
Citrate synthase	Ga0131788_12492	<i>citrate synthase</i>	Ga0248264_112029	<i>citrate synthase</i>
Citrate synthase	Ga0131788_111178	<i>citrate synthase</i>	Ga0248264_113730	<i>citrate synthase</i>
Aconitase	Ga0131788_13107	<i>aconitase</i>	Ga0248264_111436	<i>aconitase</i>
Aconitase	Ga0131788_1222	<i>aconitase</i>	Ga0248264_112069	<i>aconitase</i>
Isocitrate lyase	Ga0131788_11293	<i>isocitrate lyase</i>	Ga0248264_112494	<i>isocitrate lyase</i>
Isocitrate lyase	Ga0131788_1194	<i>cytochrome c553</i>	N/A	N/A
Malate synthase	Ga0131788_11298	<i>malate synthase</i>	Ga0248264_11809	<i>malate synthase</i>
Isocitrate dehydrogenase (NADP)	Ga0131788_13268	<i>isocitrate dehydrogenase (NADP)</i>	Ga0248264_112843	<i>isocitrate dehydrogenase (NADP)</i>

Left Branch:

Oxoglutarate ferredoxin oxidoreductase subunit A	Ga0131788_1216	<i>2-oxoglutarate ferredoxin oxidoreductase subunit A</i>	Ga0248264_111754	<i>2-oxoglutarate ferredoxin oxidoreductase subunit A</i>
Oxoglutarate ferredoxin oxidoreductase subunit B	Ga0131788_1215	<i>2-oxoglutarate ferredoxin oxidoreductase subunit B</i>	Ga0248264_111753	<i>2-oxoglutarate ferredoxin oxidoreductase subunit B</i>
Succinyl-CoA synthetase (ADP forming) subunit A	Ga0131788_111629	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113616	<i>succinyl-CoA synthetase (ADP-forming)</i>
Succinyl-CoA synthetase (ADP forming) subunit B	Ga0131788_111630	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113615	<i>succinyl-CoA synthetase (ADP-forming)</i>
Succinate dehydrogenase subunit A	Ga0131788_111667	<i>succinate dehydrogenase subunit A</i>	Ga0248264_111802	<i>succinate dehydrogenase subunit A</i>
Succinate dehydrogenase subunit D	Ga0131788_111670	<i>succinate dehydrogenase subunit D</i>	Ga0248264_111986	<i>succinate dehydrogenase subunit D</i>
Succinate dehydrogenase subunit D	Ga0131788_12494	<i>succinate dehydrogenase subunit B</i>	Ga0248264_112031	<i>succinate dehydrogenase subunit B</i>
Succinate dehydrogenase subunit C	Ga0131788_12495	<i>succinate dehydrogenase subunit A</i>	Ga0248264_112032	<i>succinate dehydrogenase subunit A</i>
Succinate dehydrogenase subunit D	Ga0131788_12496	<i>succinate dehydrogenase subunit D</i>	Ga0248264_112033	<i>succinate dehydrogenase subunit D</i>
Succinate dehydrogenase subunit D	Ga0131788_12497	<i>succinate dehydrogenase subunit C</i>	Ga0248264_112034	<i>succinate dehydrogenase subunit C</i>
Fumarase	Ga0131788_11335	<i>homodimeric fumarase (class I)</i>	Ga0248264_11703	<i>homodimeric fumarase (class I)</i>
Fumarase	Ga0131788_112217	<i>fumarase class II</i>	Ga0248264_111932	<i>fumarase hydratase-like protein</i>
	Ga0131788_111630	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113638	<i>succinyl-CoA synthetase (ADP-forming)</i>
	Ga0131788_111629	<i>succinyl-CoA synthetase (ADP-forming)</i>	Ga0248264_113639	<i>succinyl-CoA synthetase (ADP-forming)</i>
	Ga0131788_111668	<i>fumarate reductase iron-sulfur subunit</i>	Ga0248264_111803	<i>succinate dehydrogenase subunit B</i>

Genes/Pathways		Acc type IA UW3 (Ga0131788)		Acc type IIC UW6 (Ga0248264)	
TCA cycle (2-oxoglutarate dehydrogenase)		Locus tag		Locus tag	
	Ga0131788_12490	2-oxoglutarate dehydrogenase E2 co	0 0 0 0 0 0 0	Ga0248264_112027 2-oxoglutarate dehydrogenase E2 co	1 0 -1 -1 -1 1 1
	Ga0131788_12491	2-oxoglutarate dehydrogenase E1 co	0 -1 -1 1 0 1 1	Ga0248264_112028 2-oxoglutarate dehydrogenase E1 co	0 1 -1 0 -1 1 0
Glyoxylate cycle					
	Ga0131788_11298	malate synthase	-1 0 1 0 -1 1 -1	Ga0248264_11809 malate synthase	-1 -1 -1 -1 -1 1 1
	Ga0131788_12499	malate dehydrogenase (NAD)	-1 1 1 1 -1 -1 -1	Ga0248264_111563 malate dehydrogenase (NAD)	-1 1 1 1 -1 -1 0
	Ga0131788_12492	citrate synthase	-1 1 0 0 -1 1 1	Ga0248264_112029 citrate synthase	1 -1 -1 1 -1 1 1
	Ga0131788_111178	citrate synthase	1 1 -1 -1 -1 0 1	Ga0248264_113730 citrate synthase	1 0 0 0 -1 0 0
	Ga0131788_13107	aconitase	0 0 0 0 0 0 0	Ga0248264_111436 aconitase	0 0 0 0 0 0 0
	Ga0131788_1222	aconitase	-1 1 1 1 -1 -1 -1	Ga0248264_112069 aconitase	0 1 0 0 -1 -1 -1
	Ga0131788_11294	isocitrate lyase	-1 1 1 1 -1 -1 -1	Ga0248264_112494 isocitrate lyase	0 0 0 0 0 0 0
	Ga0131788_11293	isocitrate lyase	-1 1 1 1 -1 -1 -1	Ga0248264_112494 isocitrate lyase	0 0 0 0 0 0 0
NAD(P)H production					
	Ga0131788_11267	NAD(P)H dehydrogenase (quinone)	1 0 0 0 -1 0 0	Ga0248264_11621 NAD(P)H dehydrogenase (quinone)	0 0 0 0 0 0 0
	Ga0131788_1279	ferredoxin-NADP+ reductase	-1 -1 0 1 -1 1 1	Ga0248264_111097 ferredoxin-NADP+ reductase	-1 -1 -1 0 0 1 1
	Ga0131788_112417	N/A	1 0 -1 -1 -1 0 1	N/A	
Denitrification					
NapH	Ga0131788_12416	ferredoxin-type protein NapH	0 1 1 0 -1 0 0	N/A	N/A
NapG	Ga0131788_12417	ferredoxin-type protein NapG	0 1 1 0 -1 -1 -1	N/A	N/A
Cytochrome C553	Ga0131788_12420	cytochrome c553	-1 1 1 -1 -1 -1 -1	Ga0248264_113752 cytochrome c553	0 0 -1 0 0 0 1
Cytochrome C	Ga0131788_12421	cytochrome c	-1 1 1 0 -1 -1 -1	N/A	N/A
Nitrous oxide reductase	Ga0131788_12422	nitrous oxide reductase apoprotein	-1 1 1 -1 -1 -1 -1	N/A	N/A
NapC	Ga0131788_13199	cytochrome c-type protein NapC	0 1 0 -1 -1 0 1	N/A	N/A
NapF	Ga0131788_111915	ferredoxin-type protein NapF	1 0 0 0 0 0 0	N/A	N/A
NapH	Ga0131788_111918	ferredoxin-type protein NapH	-1 1 1 1 -1 -1 -1	N/A	N/A
NapG	Ga0131788_111919	ferredoxin-type protein NapG	-1 1 1 1 -1 -1 -1	N/A	N/A
NapA	Ga0131788_111920	periplasmic nitrate reductase subunit	-1 1 1 0 -1 -1 -1	N/A	N/A
NapD	Ga0131788_111921	periplasmic nitrate reductase chaperone	-1 1 1 -1 -1 -1 -1	N/A	N/A
NorQ	Ga0131788_111835	nitric oxide reductase NorQ protein	0 -1 -1 -1 -1 1 1	Ga0248264_113068 nitric oxide reductase NorQ protein	0 1 0 0 0 0 -1
	Ga0131788_111836	nitric oxide reductase NorD protein	1 -1 0 -1 -1 1 1	Ga0248264_113067 nitric oxide reductase NorD protein	0 1 0 0 -1 0 -1
Dissimilatory nitrite reductase	Ga0131788_112448	dissimilatory nitrite reductase (NO-forming)	1 -1 -1 -1 -1 1 1	Ga0248264_114768 dissimilatory nitrite reductase (NO-forming)	1 0 -1 0 0 0 0
Dissimilatory nitrite reductase	Ga0131788_12235	dissimilatory nitrite reductase (NO-forming)	-1 1 1 -1 -1 -1 -1	N/A	N/A
	Ga0131788_12378	nitrite reductase/ring-hydroxylating ferredoxin	0 -1 -1 0 0 0 1	Ga0248264_114025 nitrite reductase/ring-hydroxylating ferredoxin	0 0 -1 0 0 0 0
	N/A	N/A		Ga0248264_114773 Respiratory nitrate reductase	0 -1 -1 0 0 0 1
	N/A	N/A		Ga0248264_114772 Respiratory nitrate reductase	0 0 0 0 -1 0 1
	N/A	N/A		Ga0248264_114770 Respiratory nitrate reductase	0 0 0 0 0 0 0
Iron Transport					
	Ga0131788_11638	ferric iron transport protein A	-1 -1 -1 -1 1 1 1	Ga0248264_113926 ferric iron transport protein A	0 0 0 0 0 0 0
	Ga0131788_11639	ferric iron transport protein B	0 -1 -1 0 1 0 -1	Ga0248264_113927 ferric iron transport protein B	0 0 0 0 0 0 0
	Ga0131788_112072	FeS assembly protein IscX	0 -1 -1 -1 -1 1 1	Ga0248264_112010 FeS assembly protein IscX	0 0 -1 -1 0 1 1
	Ga0131788_112067	FeS assembly scaffold apoprotein IscS	-1 -1 -1 -1 -1 1 1	Ga0248264_112015 FeS assembly scaffold apoprotein IscS	-1 -1 -1 -1 0 1 1
	Ga0131788_12614	iron complex outer membrane receptor	0 0 1 0 -1 -1 -1	Ga0248264_111593 iron complex outer membrane receptor	0 0 0 0 0 0 0
	Ga0131788_111928	iron complex transport system ATP-binding protein	0 0 0 0 0 0 -1	Ga0248264_112405 iron complex transport system ATP-binding protein	0 0 0 0 0 0 0
	Ga0131788_111927	iron complex transport system permease	-1 1 -1 1 1 0 -1	Ga0248264_112406 iron complex transport system permease	0 0 0 0 0 0 0
	Ga0131788_112567	iron-binding CDGSH zinc finger protein	0 -1 -1 -1 -1 1 1	N/A	N/A
	Ga0131788_112068	iron-sulfur cluster assembly protein	0 -1 -1 -1 -1 1 1	Ga0248264_112014 iron-sulfur cluster assembly protein	-1 -1 -1 -1 1 1 1
	Ga0131788_1385	iron-sulfur cluster insertion protein	-1 -1 -1 -1 -1 1 1	Ga0248264_112709 iron-sulfur cluster insertion protein	-1 0 0 0 -1 0 1
	Ga0131788_12608	iron(III) transport system permease p	0 -1 -1 0 1 1 -1	Ga0248264_11461 iron(III) transport system permease p	1 0 0 1 0 1 -1
	Ga0131788_12609	iron(III) transport system substrate-binding protein	-1 -1 -1 -1 1 1 0	Ga0248264_11460 iron(III) transport system substrate-binding protein	0 -1 -1 1 1 -1 -1
	Ga0131788_12607	iron(III) transport system ATP-binding protein	0 0 0 0 0 0 0	Ga0248264_11462 iron(III) transport system ATP-binding protein	1 0 0 0 0 0 -1
	Ga0131788_111730	iron complex transport system substrate-binding protein	0 0 0 0 0 0 0	Ga0248264_112371 iron complex transport system substrate-binding protein	0 -1 -1 -1 0 1 1

Genes/Pathways		Acc type IA UW3 (Ga0131788)				Acc type IIC UW6 (Ga0248264)			
Oxidative Stress		Locus tag				Locus tag			
	Ga0131788_111044	<i>Fe-Mn family superoxide dismutase</i>	0	-1	-1	-1	0	1	1
	Ga0131788_111084	<i>thioredoxin</i>	1	0	-1	-1	-1	0	1
	Ga0131788_111140	<i>peroxiredoxin</i>	0	-1	-1	-1	0	1	1
	Ga0131788_111794	<i>AhpD family alkylhydroperoxidase</i>	-1	1	-1	-1	0	1	0
	Ga0131788_11187	<i>Cu-Zn family superoxide dismutase</i>	1	-1	-1	0	0	0	0
	Ga0131788_111880	<i>glutathione S-transferase</i>	0	-1	-1	-1	0	0	1
	Ga0131788_112206	<i>thiol-disulfide isomerase/thioredoxin</i>	0	0	0	1	0	0	-1
	Ga0131788_112392	<i>glutathione S-transferase</i>	1	0	0	0	-1	0	0
	Ga0131788_11412	<i>peroxiredoxin</i>	1	1	-1	-1	-1	-1	1
	Ga0131788_11438	<i>AhpC/TSA family protein</i>	1	0	-1	1	0	0	-1
	Ga0131788_11629	<i>glutathione S-transferase</i>	1	0	0	0	-1	0	0
	Ga0131788_11857	<i>peroxiredoxin (alkyl hydroperoxide re</i>	1	0	-1	-1	-1	1	1
	Ga0131788_11858	<i>alkyl hydroperoxide reductase subun</i>	0	0	0	0	0	0	0
	Ga0131788_12113	<i>thioredoxin reductase (NADPH)</i>	1	0	0	0	0	0	0
	Ga0131788_12189	<i>thioredoxin 1</i>	1	-1	0	1	0	-1	0
	Ga0131788_12272	<i>SOS-response transcriptional repres</i>	1	1	0	0	-1	-1	-1
	Ga0131788_12273	<i>peroxiredoxin Q/BCP</i>	1	-1	-1	0	0	1	1
	Ga0131788_12361	<i>thioredoxin reductase (NADPH)</i>	-1	1	1	-1	-1	1	0
	Ga0131788_12400	<i>thioredoxin</i>	-1	0	0	-1	-1	1	1
	Ga0131788_12412	<i>peroxiredoxin</i>	0	-1	0	1	0	0	0
	Ga0131788_12472	<i>glutathione S-transferase</i>	1	0	0	-1	-1	1	1
	Ga0131788_12507	<i>glutathione S-transferase</i>	0	0	0	-1	-1	0	0
	Ga0131788_13246	<i>glutathione synthase</i>	1	0	0	0	0	-1	0
	Ga0131788_14124	<i>glutathione peroxidase</i>	1	0	-1	0	-1	0	1
	Ga0131788_14133	<i>alkyl hydroperoxide reductase subun</i>	1	-1	-1	-1	-1	1	1
	Ga0131788_1463	<i>peroxiredoxin Q/BCP</i>	1	0	-1	-1	0	1	1
	Ga0131788_1475	<i>thioredoxin</i>	1	0	-1	1	0	1	0
	N/A	N/A							
	N/A	N/A							
	N/A	N/A							
	Ga0248264_11787	<i>Fe-Mn family superoxide dismutase</i>	0	-1	-1	1	-1	1	1
	Ga0248264_11919	<i>thioredoxin</i>	0	-1	-1	0	-1	0	1
	N/A	N/A							
	N/A	N/A							
	N/A	N/A							
	Ga0248264_114199	<i>peroxiredoxin</i>	0	0	-1	0	0	1	1
	Ga0248264_11836	<i>glutathione S-transferase</i>	0	1	0	-1	-1	0	0
	N/A	N/A							
	Ga0248264_11718	<i>AhpC/TSA family protein</i>	1	1	-1	0	-1	-1	-1
	N/A	N/A							
	Ga0248264_11357	<i>peroxiredoxin (alkyl hydroperoxide re</i>	1	-1	-1	-1	1	1	1
	Ga0248264_11356	<i>alkyl hydroperoxide reductase subun</i>	0	0	-1	0	0	1	0
	Ga0248264_112131	<i>thioredoxin reductase (NADPH)</i>	0	0	0	0	0	0	0
	Ga0248264_112845	<i>thioredoxin 1</i>	0	0	0	0	0	0	1
	Ga0248264_112358	<i>SOS-response transcriptional repres</i>	-1	0	1	1	-1	0	0
	Ga0248264_111910	<i>peroxiredoxin Q/BCP</i>	1	0	-1	0	-1	0	1
	Ga0248264_112836	<i>glutaredoxin-like protein DUF836</i>	0	0	1	0	-1	0	0
	N/A	N/A							
	N/A	N/A							
	N/A	N/A							
	Ga0248264_113575	<i>glutathione synthase</i>	1	0	1	1	-1	-1	-1
	Ga0248264_112547	<i>glutathione peroxidase</i>	1	0	0	0	-1	0	0
	Ga0248264_11989	<i>alkyl hydroperoxide reductase subun</i>	1	-1	-1	-1	0	1	1
	N/A	N/A							
	Ga0248264_111239	<i>thioredoxin</i>	-1	1	-1	1	1	0	0
	Ga0248264_114046	<i>glutathione S-transferase/GST-like p</i>	-1	1	1	-1	-1	-1	-1
	Ga0248264_112380	<i>catalase-peroxidase</i>	1	0	0	1	0	0	-1
	Ga0248264_111894	<i>cytochrome c peroxidase</i>	1	-1	-1	1	1	0	1
Other									
	Ga0131788_1337	<i>pyruvate dehydrogenase E1 compon</i>	1	-1	-1	-1	-1	1	1
	Ga0131788_1336	<i>pyruvate dehydrogenase E2 compon</i>	1	-1	-1	-1	-1	1	1
	Ga0131788_112403	<i>pyruvate ferredoxin oxidoreductase</i>	1	0	-1	-1	-1	0	0
	Ga0131788_112402	<i>pyruvate ferredoxin oxidoreductase</i>	1	1	0	-1	-1	0	0
	Ga0131788_112404	<i>pyruvate ferredoxin oxidoreductase</i>	1	1	-1	0	-1	0	0
	Ga0131788_1474	<i>ferredoxin</i>	1	1	-1	1	-1	-1	-1
	Ga0131788_112417	<i>ferredoxin-NADP+ reductase</i>	1	0	-1	-1	-1	0	1
	Ga0131788_1279	<i>ferredoxin-NADP+ reductase</i>	-1	-1	0	1	-1	1	1
	Ga0131788_111830	<i>pyruvate kinase</i>	1	0	0	1	-1	0	0
	Ga0131788_12331	<i>pyruvate kinase</i>	0	0	-1	1	0	0	0
	Ga0131788_12114	<i>CRP/FNR family cyclic AMP-depend</i>	0	1	-1	1	-1	0	0
	Ga0131788_111824	<i>CRP/FNR family transcriptional regul</i>	1	0	0	0	-1	-1	0
	Ga0131788_12259	<i>CRP/FNR family transcriptional regul</i>	1	-1	-1	-1	-1	0	1
	Ga0131788_111015	<i>phasin family protein</i>	1	1	1	-1	-1	-1	1
	Ga0131788_111147	<i>phasin family protein</i>	-1	1	1	1	-1	-1	-1
	Ga0131788_111148	<i>phasin family protein</i>	-1	1	1	1	-1	-1	-1
	Ga0131788_112134	<i>phasin protein</i>	1	0	0	0	-1	0	0
	Ga0131788_12162	<i>phasin protein</i>	1	-1	-1	-1	-1	0	1
	Ga0248264_112673	<i>pyruvate dehydrogenase E1 compon</i>	0	0	0	0	0	0	0
	Ga0248264_112672	<i>pyruvate dehydrogenase E2 compon</i>	0	0	0	0	0	0	0
	Ga0248264_111877	<i>pyruvate ferredoxin oxidoreductase</i>	1	0	0	1	-1	0	-1
	Ga0248264_111876	<i>pyruvate ferredoxin oxidoreductase</i>	1	-1	0	0	0	0	0
	Ga0248264_111878	<i>pyruvate ferredoxin oxidoreductase</i>	1	0	0	1	-1	0	-1
	Ga0248264_111240	<i>ferredoxin</i>	1	-1	-1	1	0	0	0
	N/A	N/A							
	Ga0248264_111097	<i>ferredoxin-NADP+ reductase</i>	-1	-1	-1	0	0	1	1
	Ga0248264_113073	<i>pyruvate kinase</i>	1	0	-1	-1	-1	-1	1
	Ga0248264_113202	<i>pyruvate kinase</i>	0	0	0	0	0	0	0
	Ga0248264_112132	<i>CRP/FNR family cyclic AMP-depend</i>	0	0	0	0	-1	0	0
	N/A	N/A							
	Ga0248264_112325	<i>CRP/FNR family transcriptional regul</i>	0	0	0	0	0	0	0
	Ga0248264_111898	<i>phasin family protein</i>	0	0	0	0	0	0	0
	Ga0248264_113667	<i>phasin family protein</i>	0	0	-1	0	0	1	0
	Ga0248264_113668	<i>phasin family protein</i>	0	1	0	0	-1	0	0
	N/A	N/A							
	N/A	N/A							

Intentionally blank page

Chapter 4: iCAP366: A genome-scale metabolic network reconstruction of the archetypal PAO *Accumulibacter phosphatis*

Matthew Kizaric [§], Francisco Moya-Flores [§], Brandon Nutley, Joshua H. Hamilton
and Katherine D. McMahon.

[§] These authors contributed equally to this work.

Experiment design: FMF, MK, JJH and KDM

Data analysis: FMF, MK and BN

Manuscript writing: FMF, MK, and KDM

Abstract

Self-assembled communities that simulate activated sludge performing enhanced biological phosphorus removal (EBPR) have been largely used as a model ecosystem to understand microbial interactions and process performance. This ecosystem is enriched in the uncultured Polyphosphate Accumulating Organism *Candidatus Accumulibacter phosphatis*. This microorganism constantly adapts its global physiological response across biphasic cycles of feast and famine conditions within a diverse microbial community, by simultaneously cycling polyphosphate, polyhydroxyalkanoates, and glycogen. While metabolic models have become a powerful tool to connect *Accumulibacter*'s genotype to its phenotype, progress towards metabolic modeling of this microorganism has been slow due to several challenges such as the integration of large-scale genomic data and its interpretation. In this study, we present iCAP366, a manually curated genome-scale metabolic reconstruction for *Accumulibacter phosphatis* clade IIA-UW1. The reconstruction includes 458 genes, covering 14.23% of the protein-coding genes with function prediction in the genome. Here, Flux Balance Analysis was used to predict metabolic fluxes at different glycogen reserves to acetate uptake ratios under anaerobic conditions. iCAP366 was also used to predict metabolic flux distributions through key pathways including CO₂ fixation, reductive TCA cycle, and H₂ production. Overall, iCAP366 shows qualitative and quantitative agreement with experimental observations. Thus, iCAP366 provides concepts and a basis for extensive future studies of this bacterium and other related bacteria.

Main

The biotechnological potential of a biological system can be inferred from its metagenome, and can be represented as genome-scale metabolic network reconstructions that serve as a knowledge base of all the biochemical information about an organism. Integrating this information in a structured fashion has enabled its translation into computational models that can be used to calculate metabolic phenotypes. Reconstruction of metabolic models is an iterative and labor intensive process, which involves constructing a draft model with a preliminary set of reactions, genes, metabolites and constraints. The metabolic network reconstruction process is at an advanced stage of development and has been translated into a 96-step standard operating procedure ¹¹¹. Briefly, genome-scale metabolic networks contain curated and systematized information about all known biochemical metabolites and reactions of a cell's metabolism encoded on its genome and described in experimental literature. The stoichiometric matrix (**S**) is a mathematical description of a genome-scale metabolic network, where each column corresponds to a metabolite and each row corresponds to a reaction. This matrix can be queried by many available modeling methods. These models and the calculated reaction fluxes are typically studied under a steady-state assumption. A steady-state mass balance constraint is imposed by the equation $[S_v = 0]$, where v is a vector containing the fluxes through each reaction in the network.

Constraint based models can address a quantitative and mechanistic genotype-phenotype relationship. Constraint-based reconstruction and analysis (COBRA)

methods help us analyze the allowable phenotypic states on a genome-scale and depend on the generation of the genome-scale network reconstruction and the subsequent application of constraints to these reconstructions to form the corresponding genome-scale model (GEM) *in silico* ¹¹². The COBRA methods rely on optimization methods (such as linear-quadratic programming, mixed integer linear programming and non-linear programming) that have been developed and are available as accessible software.

Accumulibacter

Activated sludge wastewater treatment processes are ubiquitous for the removal of organic matter and nutrients from municipal and industrial wastewaters. These systems employ uncultivated microbial communities to efficiently remediate wastewater streams prior to being discharged to the aquatic environment. Enhanced Biological Phosphorus Removal (EBPR), a variant of the activated sludge wastewater treatment process, is a biochemically complex process achieved through enrichment of a group of bacteria known as Polyphosphate Accumulating Organisms (PAO) employing alternating anaerobic and aerobic conditions ^{4,6}. In the EBPR process, the dominant organism is a member of the Betaproteobacteria in the Rhodocyclus group, named *Candidatus Accumulibacter phosphatis* ^{6,7}. No pure culture of this organism is yet available, though culture-independent molecular techniques provide much of the essential information that traditionally has been obtained using pure cultures ^{5,8,15,25,42–45}.

Lack of consensus on Phenotypic Potentials

Though *Accumulibacter* has been studied for years, there is still a lack of consensus on the phenotypic potentials that the organism can express ⁴. One debated topic is whether *Accumulibacter* can perform full TCA cycling anaerobically or is limited to a split reductive TCA. Several metagenomics, metatranscriptomics and proteomics studies have attempted to explore this area of EBPR research; however, a consensus on the issue has yet to be resolved ^{5,44,72,103,113}. Martin, et. al ⁹ suggested genomic evidence for a novel cytochrome that would allow for full TCA operation anaerobically. Proteomic data ¹⁶ has contradicted this claim though, and many previous models of *Accumulibacter* only consider a reductive TCA cycle ^{112,114}. However, metatranscriptomics data has suggested a large expression of both Fumarate Reductase and Succinate Dehydrogenase ^{44,55}. Another more recent discovery was hydrogen production associated with anaerobic acetate uptake ⁴⁴. As there are only limited studies done on this phenomenon, it has not been widely accepted, but it could explain how *Accumulibacter* is able to balance redox in an anaerobic environment. Also, briefly mentioned in ⁴⁴ is the activation of genes involved in carbon fixation via the Calvin Cycle. Carbon fixation through anaplerotic routes are accepted in most EBPR literature, due to an activation of the glyoxylate cycle, but CO₂ fixation through the Calvin Cycle could provide another route to replacing lost carbon from oxidation of glucose and acetate.

Current State of EBPR Modeling

Many attempts have been used to model the EBPR phenotype in a quantitative way. A simplistic stoichiometric model that incorporates minimal biochemical transformations and community dynamics to simulate a EBPR system was early theorized ¹¹⁴. Flux Balance Analysis (FBA) models have also been proposed ^{115,116}. Pramanik, et al. used hypothesized pathways with little to no genomic evidence to develop a robust FBA model ¹¹⁵ and Silva, et al. simplified the carbon transformations and electron balancing of TCA and glycolysis to obtain an accurate albeit simple model ¹¹⁶. Due to a wider availability of high quality *Accumulibacter* genomes and multi-omics datasets, a genome scale model has the potential to provide a way to tie recent advances in EBPR research together and act as a tool to study previously unexplored areas of a diverse *Accumulibacter* physiology.

Motivation

We employed omics data integration as constraints to create, validate, calibrate and improve the predictive powers of metabolic models of *Accumulibacter*. Here we present iCAP366, a manually curated genome-scale metabolic reconstruction for *Candidatus Accumulibacter phosphatis* strain UW1 consisting of 366 transformation reactions (100 reversible and 266 irreversible) and 7 transport reactions. The reconstruction includes 458 genes, covering 14.23% of the protein coding genes with function prediction in the genome. Here, FBA was used to predict metabolic fluxes at different glycogen reserves to acetate uptake ratios under anaerobic conditions. iCAP366 was also used to predict metabolic flux distributions

through key pathways including CO₂ fixation, reductive TCA cycle and H₂ production. Overall, iCAP366 shows good qualitative and quantitative agreement with experimental observations. Thus, iCAP366 provides concepts and a basis for extensive future studies of this bacterium and other related bacteria.

Methods

4.1 Computational refinement

A tier 3 Biocyc Pathway/Genome Database (PGDB) was generated using Pathway Tools software version 18.5 from the RefSeq annotation of *Candidatus Accumulibacter phosphatis* clade IIA str. UW-1 (NCBI:txid522306). Pathways were scored using Pathologic⁹¹ with a default cutoff score of 0.15 and taxonomic pruning enabled. 191 known pathways were selected, 37 of 42 new pathways were added, 30 of 43 were pruned after failing the scoring. Gap filling was then done, and 18 inferred transport reactions were added. (Supplementary Data 1 - 3).

4.2 COBRApy Model Creation

The PGDB was imported as a COBRApy¹¹⁷ model. For simplicity, only two cellular compartments were defined; compartment "e", the extracellular matrix, and compartment "c", the cytosol. Known *Accumulibacter* pathways were then selected to analyze via FBA. The pathways selected incorporated central carbon metabolism through the EMP and TCA cycle and polyhydroxyalkanoate (PHA) production. For amino acids, nucleic acids, and cofactor metabolism, pathways were added based on genome-scale prototrophies. The Calvin cycle was added after the genome displayed a nearly full carbon fixation pathway through RuBisco^{5,44}. Lipid catabolism and anabolism reactions were added to assist in the linear solving methods. Further reactions that had strong transcriptional and literary evidence supporting their existence were added as well. Variants of traditional *Accumulibacter* pathways were

added if their genes were contained in the annotated genome. Pathways that exist in similar *Accumulibacter* clades such as denitrification were added to the model; however, they were constrained to be inactive for the core phenotypic model.

4.3 Defining Anaerobic Constraints and Objectives

A standard set of constraints for anaerobic *Accumulibacter* physiology was created so the effects of more diverse phenotypic potentials could be compared. Using the model provided by Silva, et al.¹¹⁶, the general genome scale model from above was constrained to conduct a reductive TCA cycle, no hydrogen production, and no Calvin cycle activity. Other constraints were added to fully represent anaerobic conditions, prevent cycles or loops in the solution, and to help bound the solution to a reasonable space. The three aforementioned reactions were chosen to represent hypothesized phenotypes that have little or conflicting omics data. A set of simulations were defined by all possible combinations of these reactions (Table 3). A simple objective was picked where the model attempts to maximize PHA in the form of both PHB (Polyhydroxybutyrate) and PHV (Polyhydroxyvalerate) production.

4.4 Simulation of the Acetate and Glycogen flux ratio

To better understand the repercussions of modifying certain reactions during acetate uptake, the fraction of carbon flux from acetate and glycogen was fixed over ratios from 0 to 1. This was done to simulate how various *Accumulibacter* phenotypes that consume acetate and glycogen at different rates yield different

PHA production results. During each iteration, the total carbon flux from acetate and glycogen was kept constant. Flux balance analysis was run with these fixed carbon ratios to estimate the fluxes of each reaction in the optimal solution. This was done for each different hypothesized phenotype using the modified constraints from above. PHA and Pi fluxes were then graphed and compared with the results from the standard model.

4.5 Comparison of Measured and Simulated Data

For each simulation, the maximum fluxes of PHB or PHV production were recorded for each ratio of carbon flux. Though not explicitly tested, the area below this max flux line was assumed to be the feasible solution space of the simulation. PHA data from several EBPR studies that were aggregated by Silva, et al.¹¹⁶ were normalized by total carbon flux coming from Acetate feed and intracellular glycogen reserves. Unlike the referenced study, no reconciliation was done during our simulations, as our model could have fixed carbon dioxide or balanced redox using mechanisms not considered in Silva, et al.¹¹⁴. This could have increased the feasible solution space to capture more data points without the need for reconciliation. All datasets that came from PAO enrichments (43 out of 55) were selected and plotted against the simulated solution spaces for comparison.

4.6 Estimated PHA and Pi Stoichiometry

To provide an example for how this FBA model could be used to predict anaerobic PHA production and Pi release, an estimation of metabolite stoichiometry was

computed. First, the average and standard deviation of glycogen/acetate ratios were calculated from the measured data discussed above. The fluxes of PHB, PHV production, glycogen utilization, acetate uptake and Pi release were then considered for each proposed phenotype at a glycogen/acetate ratio equaling the measured dataset's average. For error reconciliation, fluxes for the polymers and Pi were also recorded at the average plus one standard deviation, and the average minus one standard deviation. The time taken to uptake one mmol of acetate was computed from the acetate uptake rate (assuming a constant uptake rate), was used to calculate the concentration change of PHA species and Pi per unit of acetate consumed. This was repeated for the upper and lower error bounds and across all simulations.

Results

EBPR is a complex dynamic process, and many kinetic parameters are required for solution of a dynamic model. Here, the anaerobic EBPR phase was considered independently using a flux-based steady-state model. In this framework, metabolite fluxes were constrained by mass conservation, thermodynamics (reaction directionality), and an assumption of pseudo steady state. More specific to the EBPR process, by controlling the source of carbon flux into the *Accumulibacter* metabolism, poly-P and PHA synthesis/degradation, acetate uptake, Pi release and CO₂ fixation were all able to be critically examined and compared with *in-vivo* experiments.

Creation of a standard anaerobic model

A set of constraints and objective functions were created to closely resemble the known phenotypic response of *Accumulibacter* under the anaerobic phase of an EBPR cycle. Numerous studies have shown that PHA synthesis occurs during this phase, as well as phosphorus release in the system. As *Accumulibacter* is mainly acetate-fed under laboratory conditions, PHB is assumed to be the dominant PHA synthesized, thus our standard model attempts to maximize PHB production when acetate is present. Additionally, it is assumed that *Accumulibacter* will minimize carbon dioxide release and glycogen degradation as the latter represents wasted carbon resources and the former would lead to a reduction in valuable internal carbon reserves. The constraints of the standard model represent an acetate-fed, non-denitrifying, anaerobic phase with no net growth for the organism at a constant pH. Certain phenotypes that are not well characterized, such as the Calvin Cycle, full TCA activity via succinate dehydrogenase (SDH), and hydrogen production were also constrained to be initially inactive. However, in subsequent simulations some of these constraints were changed to test the feasibility and change in PHA production that occurs when the phenotypes are active.

Simulation of Anaerobic Carbon Utilization

Figure 1 corroborates the results from the standard model described in the previous FBA study ¹¹⁶. From a carbon source ratio of 0 to 0.3386, the model predicts generation of reducing power (in the form of NADH) from the glyoxylate cycle, thus the PHA production is a mix of both PHV and PHB. At the maximum PHA flux, all

PHA production is in the form of PHB, as the reducing power generated from glycolysis equals the required NADH for PHB creation. From a ratio of 0.339 to 0.737, the model increases PHV flux and the reductive TCA cycle is used to balance redox. After a ratio of 0.737, a constrained CO_2 fixation via PEP-Carboxylase causes PHV flux to stabilize and thus total PHA flux to decrease.

Figure 2 shows the fluxes of PHA production and P_i release per unit of Cmmol utilization (1 Cmmol coming from a mix of glycogen and acetate) for each of the 8 simulations. Similar phases were observed here, however, the activation of other pathways allowed for differences in PHA fluxes. This also depicts the max flux achieved for each of the simulations, and the glycogen to acetate ratio where the maximum was observed. In the FBA runs where the Calvin Cycle was activated, the decline in PHB production in the simulations seemed to shift to the right (more of the total carbon originating from glycogen). In addition, a near maximum PHB flux was able to be achieved for a larger portion of the simulation when the Calvin Cycle was active. The simulations with an active TCA cycle seemed to only allow small values of PHV production with a small glycogen to acetate ratio, whereas simulations with only a reductive TCA cycle allowed for larger PHV production when most of the carbon came from acetate. Hydrogen production restricted the amount of PHV that could be created in the higher ranges of glycogen to acetate ratios. Additionally, a larger maximum value of P_i release was seen when H_2 production was allowed.

Resolution of FBA results and its validation

A total of 86 observed flux data points retrieved from Silva, et. Al ¹¹⁶ were considered for comparison to the simulation's output. Of them, the simulations with a fully active TCA cycle had solution spaces in which most of the measured PHB data could fit; however the solution space of these simulations for measured PHV contained barely any measured data (Figure 3). Of the simulations, the standard model with hydrogen production (simulation 6) captured the greatest number of data points for PHA production under its solution space (56 out of 86 total data points being within its potential space). Both simulations 1 and 3 had the lowest number of points due to no PHV data points within the viable area. Other simulations had similar results ranging in at nearly half of total data points measured.

Estimation of stoichiometric constants

Figure 4 shows the estimates of total PHA's produced and Pi released given one Cmmol of acetate uptaken for each of the eight conditions explored. Additionally, the estimated amount of glycogen to be degraded was calculated to be 0.253 Cmmol, with an error ranging from 0.368 to 0.0575 across all simulations (data not shown). Both PHB creation and Pi release had similar stoichiometric constants among all simulations. However, for PHB there was a smaller range of error between the reductive TCA cycle compared to the FBA runs with a full TCA cycle. PHV was inconsistent between most scenarios, with some predicting no PHV production and increased error ranges (Figure 4).

Discussion

With the large influx of high-throughput omics studies in the EBPR research, there are several hypothesized phenotypic potentials under active discussion. However, it is still somehow difficult for metabolic and kinetic experimentation to understand results from whole microbiome studies. Thus a model can provide a framework for researchers to test how different constraints on an EBPR system can result in a wide variety of observed data. As all the constraints discussed in our work have genomic evidence, this study was intended to suggest what types of kinetic and metabolic data would support the inclusion of these reactions into a standard EBPR metabolic model. Varying the ratio of carbon flux from glycogen and acetate was done to simulate the wide variety of kinetic data surrounding glycogen degradation, acetate uptake, and PHA production. By constraining several key reactions, the model was able to show how a handful of reactions can change the expressed phenotype drastically. Furthermore, these simulations can help find kinetic points of interest, such as maximum PHA production, maximum Pi release, and fluxes where certain phenotypes are not expressed.

Lack of Evidence for an Active Calvin Cycle During Acetate Uptake

Though some studies acknowledge the presence of carbon fixation genes in *Accumulibacter* ^{5,44}, very little kinetic studies have explored the implication of these genes being active. Our model predicted carbon fixation through the Calvin Cycle immediately following the maximum points of PHA production. This suggests that if the Calvin Cycle were to be active during acetate uptake, there would a large range

of acetate uptake rates that gave near optimal PHB synthesis. Looking into the metabolic reasoning behind this, is the fact that at carbon flux ratios larger than the maximum point of PHA production, a reductive TCA cycle is active in all models. This allows for extra reducing potential to be used in PHV formation to help balance redox. In the models where the Calvin cycle is active, the extra CO₂ fixed provides more carbon for reduction into PHB, delaying the need for a reductive TCA until Rubisco reached an imposed upper bound of 0.1 flux (Table 2, Supplementary data 1).

Though the biology behind this is interesting to consider, when compared to the aggregated data from *in-vivo* studies¹¹⁶, there is little evidence to support that this occurs. From Figure 3, the simulations with an active Calvin Cycle had a large area in the mid-range of ratios tested where no PHV was created. This decrease in PHV flux directly contradicted the measured PHV where most values were in this “valley”. Other evidence can be found in the estimated stoichiometric values in Figure 4, with only the simulations lacking Calvin Cycle activity had a clear PHV value. These values corresponded with roughly a 5% PHV composition of total polymer, which is close to reported values in numerous studies.

Hydrogen Production Expands PHA Production Space

Accumulibacter encodes for both NADH- and ferredoxin-dependent hydrogenases⁵. In the model, we found that they had distinct flux profiles; where the NADH-dependent hydrogenase was expressed at high ratios, whereas the ferredoxin-dependent variant was active when the major carbon source was acetate (data not

shown). NADH-dependent hydrogen production caused PHB to have greater flux values in the high ratio ranges for each scenario analyzed. This is likely due to better redox balancing leading to an increased PHB flux during high glycolysis flux.

In the ferredoxin-dependent case, the ferredoxins are reduced by a 2-oxoglutarate ferredoxin oxidoreductase, an enzyme in the TCA cycle that replaces 2-oxoglutarate dehydrogenase. This helps balance redox by removing one reaction in the TCA cycle that produces NADH. In the full TCA cycle, this leads to a greater PHB flux since less acetate needs to be converted to PHV to balance redox. In the reductive TCA examples, the maximum flux of PHV synthesis occurs at lower Glycogen/Acetate flux ratios, because 2-oxoglutarate ferredoxin allows for a balanced path through the lower part of the TCA cycle to propynyl-CoA. The total solution space of PHA production is increased in both cases.

A Complete Anaerobic TCA Cycle Could Be a Feasible Phenotype

It has been a well debated and hypothesized topic that *Accumulibacter* may be able to have a full anaerobic TCA cycle, but generally there is no consensus on how the quinol from succinate dehydrogenase could be re-oxidized. Our FBA approach did observe a complete anaerobic TCA activation in ranges varying from 0 to 0.498 glycogen to acetate ratios in simulations 1 and 3, and from 0 to 0.3286 in simulations 2 and 4. At these points, ATPase flips from creating a proton motive force (*pmf*) to using a *pmf* for ATP synthesis (data not shown). The creation of the proton motive force could allow for a NADH or NADPH to oxidize the quinols created

from the TCA cycle. The model was able to balance redox cofactors with a quinol reducing NAD at the expense of 4 protons in the *pmf*.

However, it is important to note that FBA modeling considers limited thermodynamic information. Little information is known about the ATP synthase complex and electron transport chain in *Accumulibacter* ¹¹⁶. Studies on the role of *pmf* have shown that membrane potential has a larger impact on acetate uptake than a concentration gradient ^{118,119}. Membrane potential is dependent on numerous factors, and with an organism that is known to have very active transport of ions across a membrane, *pmf* maintenance is likely a highly coordinated system *in-vivo* that cannot be modeled with FBA alone. Additionally, Burow, et al. showed evidence that inhibition of fumarate reductase complex does not affect acetate uptake though the generation of a *pmf*, but did not comment on activity of the succinate dehydrogenase's impact on *pmf* ¹¹⁸. Regardless, the observation of a feasible oxidative TCA cycle suggests that *Accumulibacter*'s physiology is quite diverse and may have many hidden phenotypes to be explored.

Future Directions

Our proposed model is meant to be used as an exploratory tool in assisting EBPR researchers to direct further endeavors. There are several immediate technical areas where improvements could be made. First and foremost, an updated constraint dictionary could drastically change *in-silico* phenotypes. For example, in this study an arbitrary ceiling of 0.1 mmol*L⁻¹*hr⁻¹*gDW⁻¹ was set for both carbon fixation reactions (RuBisCo and PEP-Carboxylase). This was done to minimize the

amount of incoming carbon flux that could come from inorganic CO₂ as previous calibration without an upper bound would lead to run-away carbon fixation. As these reactions are considered slow, the constraint was set without knowing the enzyme kinetics of *Accumulibacter*'s RuBisCo or PEP-Carboxylase. There are many reactions in the central carbon cycle that could drastically change the shape of hypothesized flux profiles, however, enzyme kinetic data is poor for *Accumulibacter*, making accurate constraints difficult to judge.

Furthermore, there are many more anaerobic conditions that could be explored, such as the use of alternative carbon sources, the potential for denitrification, cell growth, etc. All these conditions have evidence of feasibility from genomics, metatranscriptomics data, or experimental results, thus it would be interesting to see how the model behaves under varying constraints, and if it corresponds to *in-vivo* data.

Lastly, while the analysis that were done in this study focused on anaerobic phenotypes, the reactions that are all present for predicting aerobic conditions are included in the model. The FBA analysis of aerobic conditions were ignored to focus on topics that are more discussed in the field; however, understanding how phosphate is re-uptaken, glycogen synthesized, and *Accumulibacter* grows is vital to designing more resilient EBPR systems.

Acknowledgements

This work was partially supported by funding from the National Science Foundation (MCB-1518130). Additional funding from the Chilean National Commission for Scientific and Technological Research (CONICYT) as a fellowship to Francisco Moya is also acknowledged. The authors also thank Christopher Lawson and Matthew Scarborough for their support.

Figures Descriptions

Figure 1 | Simulation of *Accumulibacter*'s phenotypic response to varying glycogen to acetate flux ratios under "standard" constraints.

Allowable solution spaces for PHA production and Pi release under a reductive TCA cycle, no hydrogen production, and no Calvin cycle activity during anaerobic acetate uptake. For each point, the fraction of carbon flux from acetate and glycogen was fixed over ratios from 0 to 1.

Figure 2 | Flux profiles of proposed anaerobic mechanisms.

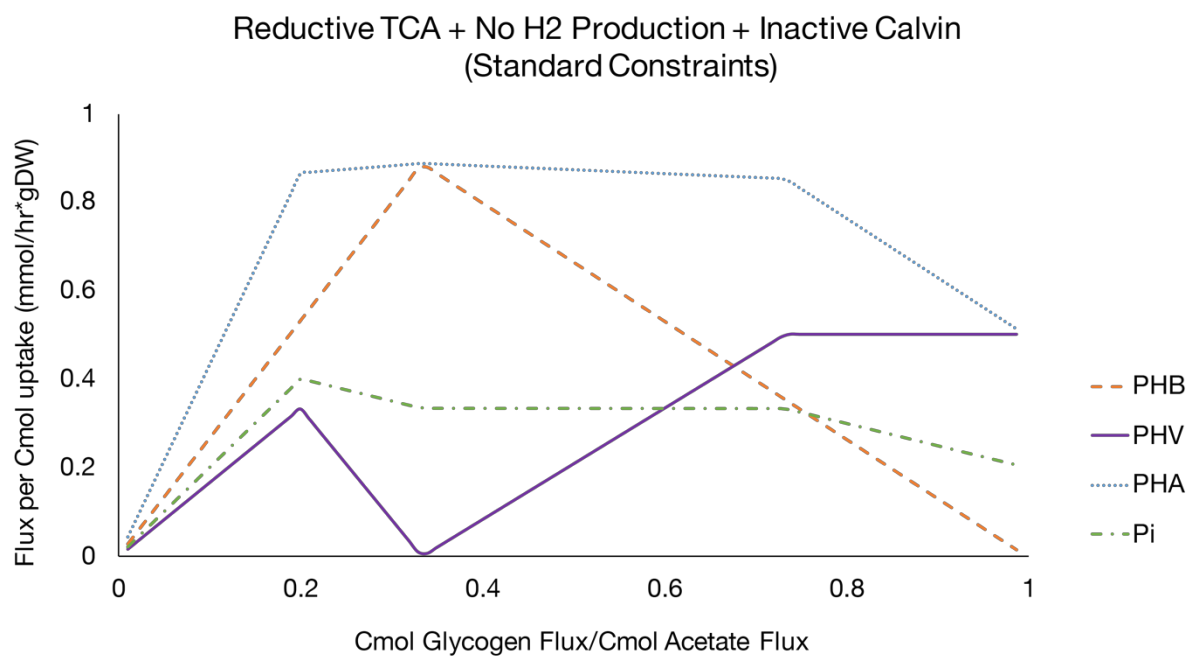
Flux profiles of PHA production and Pi release per unit of Cmmol utilization for each of the eight simulations analyzed under this study. As described above, the glycogen to acetate flux ratio was varied from 0 to 1 and the optimal solution was found under each proposed mechanism. See Table 3 for all possible scenarios defined by combining the selected constraints.

Figure 3 | Resolution of FBA simulations with previously obtained results.

For each simulation in Figure 2, the maximum fluxes of PHB or PHV production were recorded for each ratio of carbon flux. PHA data from several EBPR studies that were aggregated by Silva, et al.¹¹⁶ were normalized by total carbon flux coming from acetate feed and intracellular glycogen reserves. All datasets that came from PAO enrichments were selected and plotted against the simulated solution spaces for comparison.

Figure 4 | Model estimated PHA production and Pi release per 1 Cmmol of acetate.

Using aggregated literature data¹¹⁶, iCAP366 was able to estimate the total production of commonly studied metabolites per Cmmol of acetate uptaken for every hypothesized phenotype.

Figures**Figure 1**

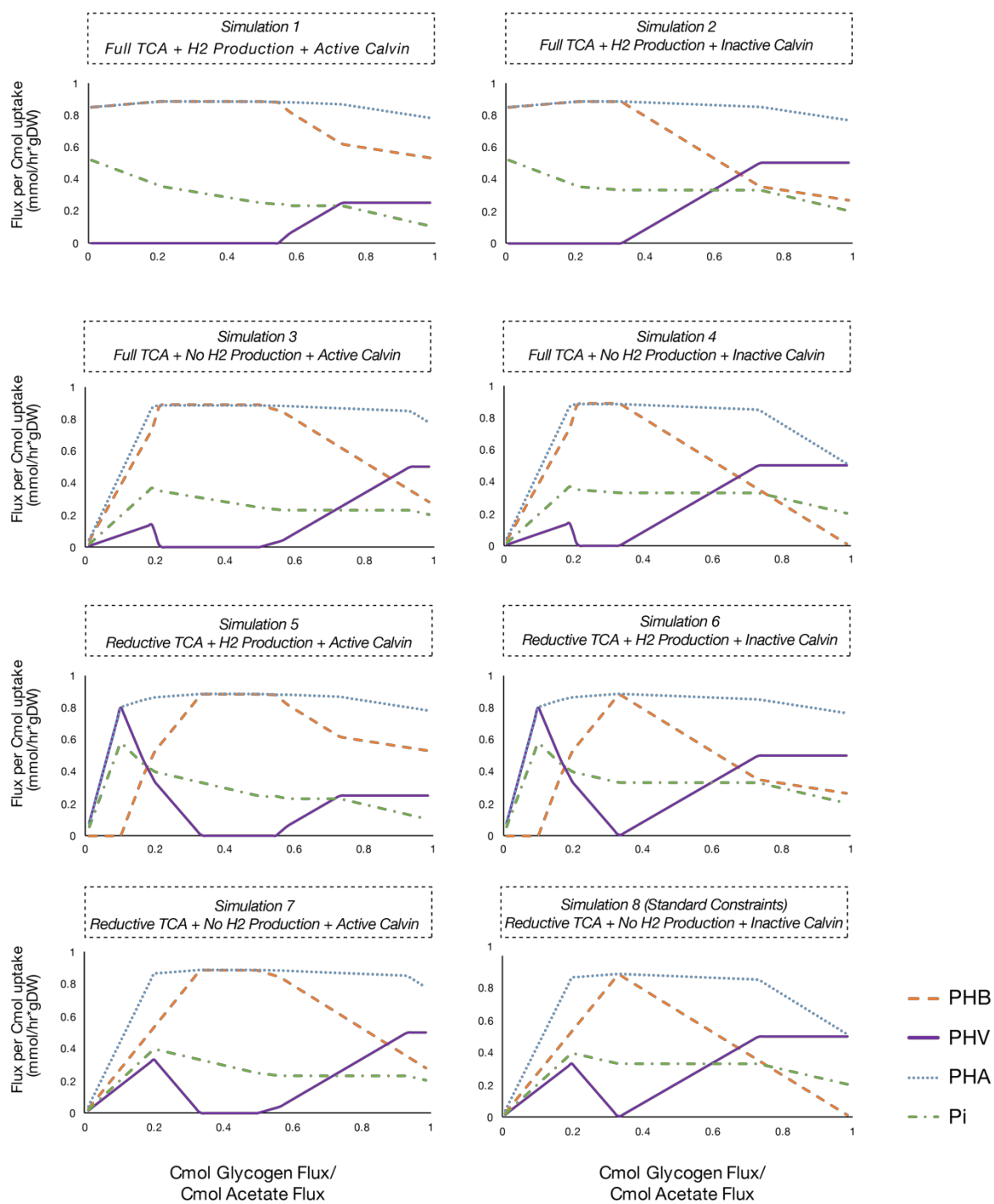


Figure 2

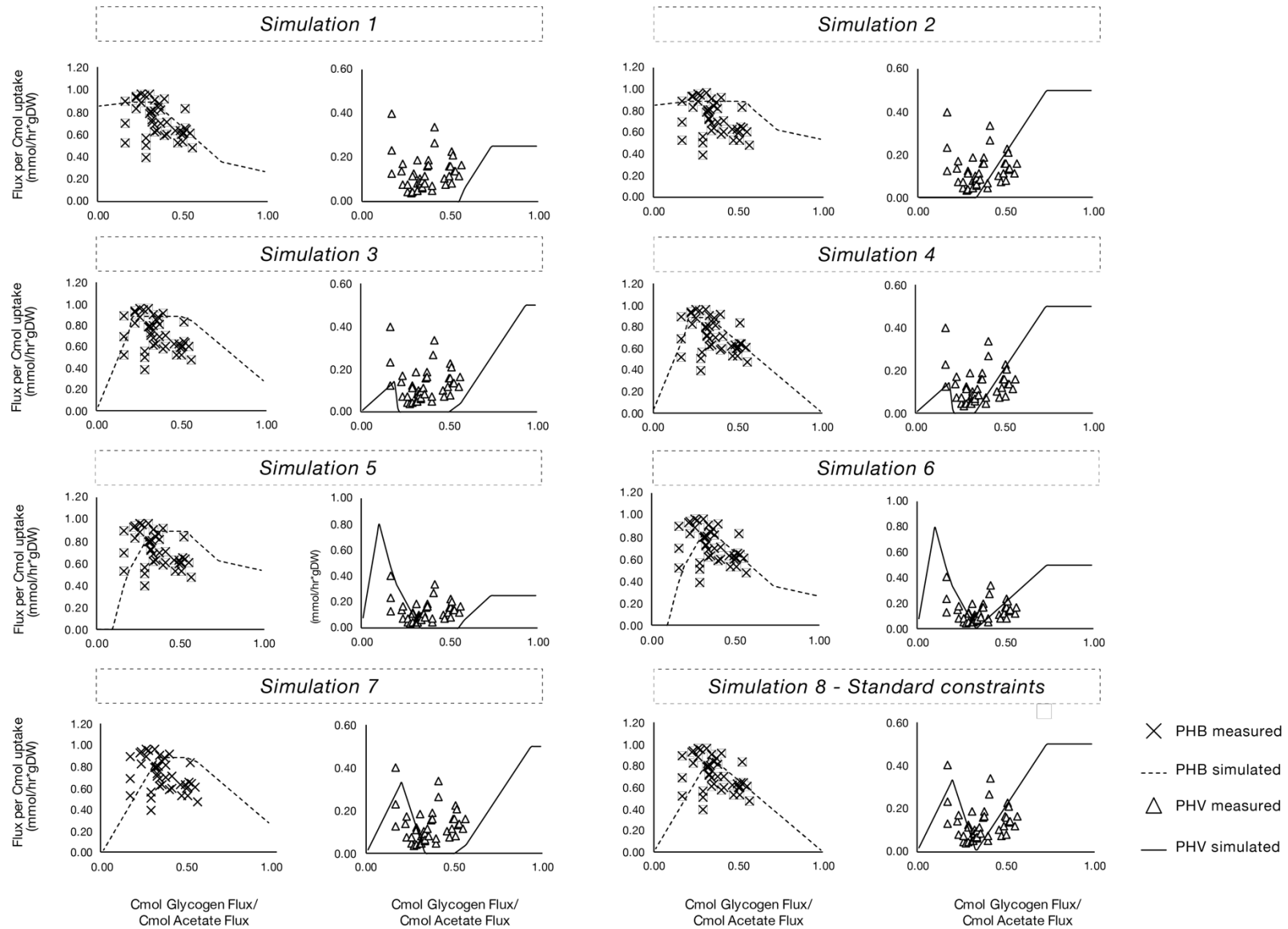
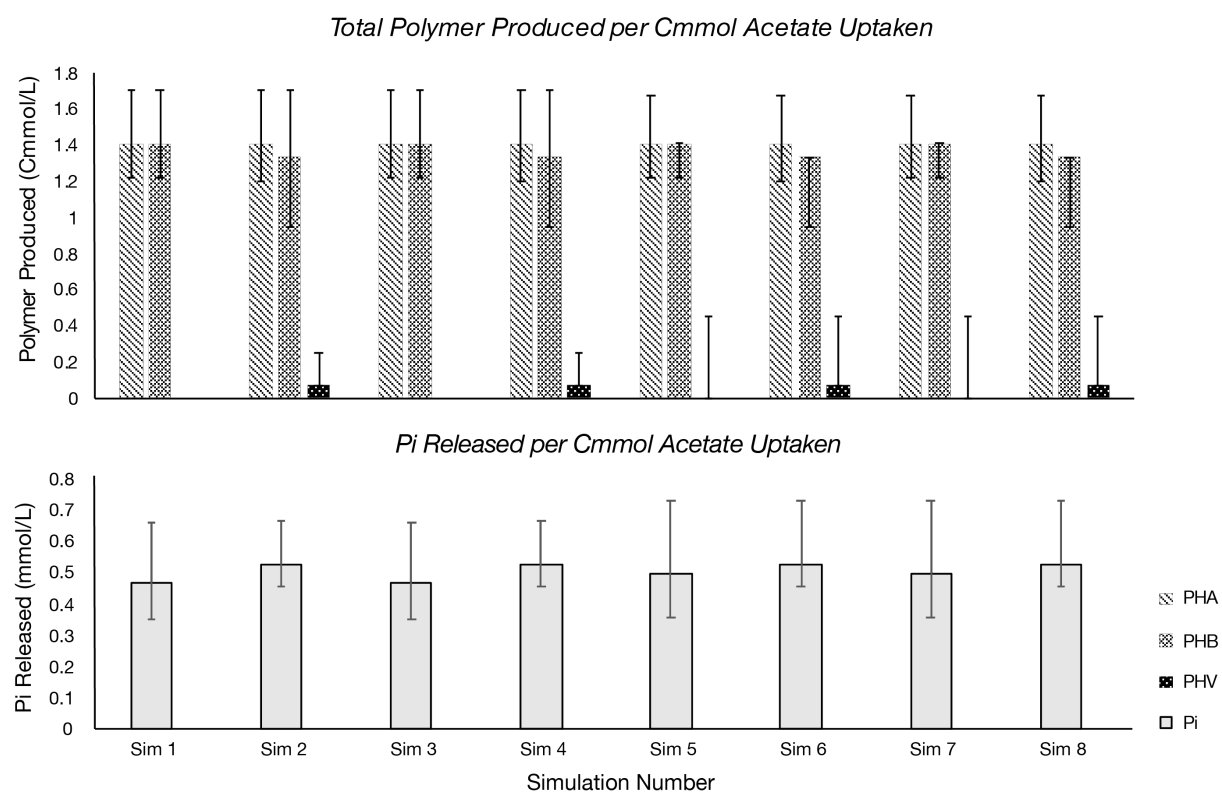


Figure 3

**Figure 4**

Tables**Table 1 |** Summary of iCAP366 GEM.

Total Reactions	366
Reversible	100
Non-reversible	266
Boundary Reactions	80
Transport Reactions	7
Mapping Reactions	2
Total Genes	458
Reactions with known genes	273
Enzymatic reactions w/o genes	5
Pathways Added	63

Table 2 | Pathways constraints used during this study.

Phenotype	Reaction	Default constraint (upper bound, lower bound)	Modified constraint
Full TCA	Fum + QH ₂ --> Succ + Q	(0, 1000)	(-1000, 1000)
H ₂ Production	H ₂ <-->	(0,0)	(0, 0.7)
Active Calvin	Ribulose-1,5-P + CO ₂ + H ₂ O --> 2 G3P + 2 H ⁺	(0,0)	(0, 0.1)

Table 3 | All possible scenarios defined by combining the selected phenotypes and selected constraints.

Simulation Number	Full TCA	H₂ Production	Active Calvin
1	Modified	Modified	Modified
2	Modified	Modified	Default
3	Modified	Default	Modified
4	Modified	Default	Default
5	Default	Modified	Modified
6	Default	Modified	Default
7	Default	Default	Modified
8	Default	Default	Default

Supplementary Material

Supplementary table S1 | Maximum Polymer and Pi fluxes obtained for each simulation.

Sim. #	Full TCA	H2 Production	Active Calvin	PHB		PHV		PHA		Pi	
				Max Value mol/(L*gDW*hr*Cmol util.)	Gly/A cet Cmol/Cmol	Max Value mol/(L*gDW*hr*Cmol util.)	Gly/A cet Cmol/Cmol	Max Value mol/(L*gDW*hr*Cmol util.)	Gly/A cet Cmol/Cmol	Max Value mol/(L*gDW*hr*Cmol util.)	Gly/A cet Cmol/Cmol
1	Modified	Modified	Modified	0.888	0.219	0.250	0.827	0.888	0.219	0.517	0.010
2	Modified	Modified	Default	0.888	0.219	0.500	0.737	0.888	0.219	0.517	0.010
3	Modified	Default	Modified	0.888	0.219	0.500	0.936	0.888	0.219	0.371	0.189
4	Modified	Default	Default	0.888	0.219	0.500	0.737	0.888	0.219	0.371	0.189
5	Default	Modified	Modified	0.888	0.339	0.797	0.100	0.888	0.339	0.578	0.100
6	Default	Modified	Default	0.880	0.339	0.797	0.100	0.887	0.339	0.578	0.100
7	Default	Default	Modified	0.888	0.339	0.500	0.936	0.888	0.339	0.398	0.199
8	Default	Default	Default	0.880	0.339	0.500	0.737	0.887	0.339	0.398	0.199

Intentionally blank page

Chapter 5: Recommendations for Future Research

Use of metatranscriptomics constraints in genome-scale modeling

Although genomics data provides a broad overview of the metabolic potential of a species, it does not convey information regarding the specificity of a metabolic process or its activation under different physiological or environmental conditions¹²⁰. To address this problem, general approaches that can suggest experimentally testable hypotheses to reconcile inconsistencies between simulation and experimental data continue to be needed. The integration of cellular processes, supported by high-throughput data types into a single mathematical model, allow us to more accurately compute complex phenotypes and will guide the discovery of unknown aspects of cellular functions beyond metabolism.

Reconstruction of metabolic models using transcriptomics data has been successfully used and tested in prokaryotic systems and has been extensively reviewed¹²¹. It considers the dynamic state of mRNA levels and has been linked to metabolism through “coupling constraints” to improve prediction accuracy. Transcriptomics also suggests regulatory mechanisms under different conditions, and differential and comparative network analysis approaches may give useful information about key regulators controlling common biological processes among organisms. Moreover, RNA-Seq has served as preferred choice as it does not need prior genome information, thus enabling transcriptome profiling for uncultivated organisms, such as *Accumulibacter*.

Analysis of the discrepancies between model predictions and experimental data allows targeted experimentation that leads to better refinement of the genotype-phenotype relationships. However, there is limited information available on the genome-wide transcriptional analysis for *Accumulibacter* under conditions different to the regular EBPR cycle and a few regulatory mechanisms involved^{80,81}. Therefore, to further build a comprehensive model that properly accounts for unanticipated constraints, it will be necessary to uncouple environmental signals that occur simultaneously during the cycle. The use of gene expression constraints under normal and perturbed conditions together would lead to a better prediction of *Accumulibacter* flux remodeling. The perturbation conditions can be designed to reflect actual disturbances or undesirable operating conditions experienced in full-scale biological nutrient removal facilities. These might include, but are not limited to:

Acetate/Oxygen contact: In the typical cycle of *Accumulibacter* cultures, exposure to acetate only occurs during anaerobic conditions. It is generally accepted that good EBPR activity requires physical/temporal separation of electron donor and acceptor, and EBPR performance deterioration in full-scale wastewater treatment plants had been associated to the presence of oxygen within the anaerobic zone¹. To our current understanding, *Accumulibacter* needs to experience bottlenecks in energy and reducing equivalent demands when taking up and storing carbon. Therefore, by adding oxygen to the medium in the middle of the anaerobic phase to study the transcriptome in response to the aerobic exposure to acetate, one could expect the oxygen to relieve those bottlenecks and disrupt the usual response.

Decoupling PHA degradation to P-uptake: A second perturbation experiment can be conducted by eliminating acetate addition during the cycle, and adding acetate in the aerobic phase of the cycle. This can help answer whether the role of the anaerobic conditions in EBPR is necessary to flux acetyl-CoA to PHA formation coupled with polyP metabolism, or just as a selective force to eliminate organisms that cannot uptake acetate anaerobically. The main difference between this experiment and the previously stated is that there would not be any fraction of the acetate metabolized in the anaerobic zone, thus we could solely investigate the transcriptional pattern of acetate conversion under aerobic conditions.

Using data obtained from these and other metatranscriptomics profiles, we could start building context-specific models for *Accumulibacter*, by identifying a flux distribution that is consistent with biological objectives and minimizing the utilization of reactions classified as inactive. We suggest using algorithms such as Gene Inactivity Moderated by Metabolism and Expression (GIMME), which uses binary gene expression data and genome-scale metabolic networks to generate context-specific reconstructions ¹²². These transcriptome datasets can be then used as experimental soft-constraints to obtain corresponding *Accumulibacter*-specific metabolic models. Combining Dynamic Flux Balance Analysis (DFBA) ¹²³ and GIMME can help to predict time-course flux profiles based on temporal gene expression patterns in the EBPR cycle.

Transcriptional Regulatory Networks

It is unlikely that genome-scale modeling alone can elucidate the physiological response of versatile bacteria such as *Accumulibacter* given any environmental condition. Indeed, *Accumulibacter* has evolved regulatory networks to integrate environmental signals or acquired differentiated states that result in modulation of gene expression⁸⁰. Therefore, it is believed that specific gene expression triggers the PAO phenotype depending on the environmental constraints. Future work should focus on comparing a larger amount of *Accumulibacter* clades transcriptional programs at different organizational levels, ranging from the co-expression patterns between genes to higher-order relationships between functional attributes. Only once deciphering the underlying transcriptional mechanisms that explain *Accumulibacter*'s behavior, implementing a combined analysis of both metabolic and regulatory networks can be applied to genome-scale models, such as the one developed under Chapter 4 of this thesis work. Systematically obtained time- and space-resolved omics datasets across multiple environmental scenarios combined with the power of metabolic modeling can allow deconvolution of structure-function relationships by identifying key genes involved in the EBPR phenotype and their functions. Such knowledge has the potential to form the foundation for discovering novel attributes of *Accumulibacter* on a much larger scale compared with previous efforts.

References cited

1. He, S., Bishop, F. I. & McMahon, K. D. Bacterial community and 'Candidatus Accumulibacter' population dynamics in laboratory-scale enhanced biological phosphorus removal reactors. *Appl. Environ. Microbiol.* **76**, 5479–87 (2010).
2. Neethling, J. B. Factors Influencing the Reliability of Enhanced Biological Phosphorus Removal. *Water Intell. Online* **5**, 9781780404479–9781780404479 (2015).
3. Oehmen, A., Yuan, Z., Blackall, L. L. & Keller, J. Comparison of acetate and propionate uptake by polyphosphate accumulating organisms and glycogen accumulating organisms. *Biotechnol. Bioeng.* **91**, 162–8 (2005).
4. Seviour, R. J., Mino, T. & Onuki, M. The microbiology of biological phosphorus removal in activated sludge systems. *FEMS Microbiol. Rev.* **27**, 99–127 (2003).
5. Martín, H. G. *et al.* Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.* **24**, 1263–9 (2006).
6. Hesselmann, R. P., Werlen, C., Hahn, D., van der Meer, J. R. & Zehnder, A. J. Enrichment, phylogenetic analysis and detection of a bacterium that performs enhanced biological phosphate removal in activated sludge. *Syst. Appl. Microbiol.* **22**, 454–65 (1999).
7. Crocetti, G. R. *et al.* Identification of Polyphosphate-Accumulating

- Organisms and Design of 16S rRNA-Directed Probes for Their Detection and Quantitation. *Appl. Environ. Microbiol.* (2000). doi:10.1128/AEM.66.3.1175-1182.2000.Updated
8. McMahon, K. D., Jenkins, D. & Keasling, J. D. Polyphosphate kinase genes from activated sludge carrying out enhanced biological phosphorus removal. *Water Sci. Technol.* **46**, 155–62 (2002).
 9. Martin, H. G., Ivanova, N., Kunin, V., Blackall, L. & McMahon, K. D. Genetic Blueprints for Enhanced Biological Phosphorus Removal (EBPR) Based on Environmental Shotgun Sequencing. *Water Environ. Found.* 82–85 (2006).
 10. He, S. & McMahon, K. D. Microbiology of 'Candidatus Accumulibacter' in activated sludge. *Microb. Biotechnol.* **4**, 603–19 (2011).
 11. Peterson, S. & Warnecke, F. Environmental distribution and population biology of Candidatus Accumulibacter, a primary agent of Biological Phosphorus Removal. *Environ. ...* **10**, 2692–2703 (2008).
 12. Flowers, J., He, S. & Yilmaz, S. Denitrification capabilities of two biological phosphorus removal sludges dominated by different "Candidatus Accumulibacter" clades. *Environ. ...* **1**, 583–588 (2009).
 13. Hambly, E. & Suttle, C. A. The virosphere, diversity, and genetic exchange within phage communities. *Current Opinion in Microbiology* **8**, 444–450 (2005).
 14. Riley, P. A. Phages: their role in bacterial pathogenesis and biotechnology. *J. Clin. Pathol.* **59**, 1003–1004 (2006).
 15. Flowers, J. J. *et al.* Comparative genomics of two 'Candidatus

- Accumulibacter' clades performing biological phosphorus removal. *ISME J.* **7**, 2301–14 (2013).
16. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45**, 1320–1328 (2000).
 17. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
 18. Abedon, S. T. *Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses. Advances in molecular and cellular microbiology*; 15 (2008).
 19. Sutherland, I. Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology* **147**, 3–9 (2001).
 20. Kunin, V. *et al.* A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res.* **18**, 293–7 (2008).
 21. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–12 (2007).
 22. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–70 (2010).
 23. Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
 24. Oyserman, B. O. *et al.* Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating

- bacteria. *ISME J.* **10**, 2931–294567 (2016).
25. Skennerton, C. T., Barr, J. J., Slater, F. R., Bond, P. L. & Tyson, G. W. Expanding our view of genomic diversity in Candidatus Accumulibacter clades. *Environ. Microbiol.* **17**, 1574–85 (2015).
 26. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–31 (2014).
 27. Kupczok, A. & Bollback, J. P. Motif depletion in bacteriophages infecting hosts with CRISPR systems. *BMC Genomics* **15**, 663 (2014).
 28. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2012).
 29. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-.).* **315**, 1709–1712 (2007).
 30. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–400 (2008).
 31. Sun, C. L. *et al.* Phage mutations in response to CRISPR diversification in a bacterial population. *Environ. Microbiol.* **15**, 463–70 (2013).
 32. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet.* **9**, e1003312 (2013).
 33. Koonin, E. V. & Wolf, Y. I. Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus–host coevolution.

- Mol. Biosyst.* **11**, 20–27 (2015).
34. Childs, L. M. *et al.* CRISPR-Induced Distributed Immunity in Microbial Populations. *PLoS One* **9**, e101710 (2014).
 35. Childs, L. M., Held, N. L., Young, M. J., Whitaker, R. J. & Weitz, J. S. MULTISCALE MODEL OF CRISPR-INDUCED COEVOLUTIONARY DYNAMICS: DIVERSIFICATION AT THE INTERFACE OF LAMARCK AND DARWIN. *Evolution (N. Y.)*. **66**, 2015–2029 (2012).
 36. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary Dynamics of the Prokaryotic Adaptive Immunity System CRISPR-Cas in an Explicit Ecological Context. *J. Bacteriol.* **195**, 3834–3844 (2013).
 37. Rho, M., Wu, Y.-W., Tang, H., Doak, T. G. & Ye, Y. Diverse CRISPRs Evolving in Human Microbiomes. *PLoS Genet.* **8**, e1002441 (2012).
 38. England, W. E., Kim, T. & Whitaker, R. J. Metapopulation Structure of CRISPR-Cas Immunity in *Pseudomonas aeruginosa* and Its Viruses. *mSystems* **3**, e00075-18 (2018).
 39. Shapiro, O. H., Kushmaro, A. & Brenner, A. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J.* **4**, 327–36 (2010).
 40. Skennerton, C. T. *et al.* Phage encoded H-NS: a potential achilles heel in the bacterial defence system. *PLoS One* **6**, e20095 (2011).
 41. Motlagh, A. M., Bhattacharjee, A. S. & Goel, R. Microbiological Study of Bacteriophage Induction in The Presence of Chemical Stress Factors in Enhanced Biological Phosphorus Removal (EBPR). *Water Res.* **81**, 1–14

- (2015).
42. He, S., Gall, D. L. & McMahon, K. D. 'Candidatus accumulibacter' population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Appl. Environ. Microbiol.* **73**, 5865–5874 (2007).
 43. Albertsen, M., Hansen, L. B. S., Saunders, A. M., Nielsen, P. H. & Nielsen, K. L. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J.* **6**, 1094–106 (2012).
 44. Oyserman, B. O., Noguera, D. R., Del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus Accumulibacter phosphatis*. *ISME J.* (2015).
doi:10.1038/ismej.2015.155
 45. Camejo, P. Y. *et al.* *Candidatus Accumulibacter phosphatis* clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors. *Water Res.* **102**, 125–37 (2016).
 46. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–63 (2011).
 47. Zerbino, D. R. Using the Velvet de novo assembler for short-read sequencing technologies. *Curr. Protoc. Bioinforma.* **Chapter 11**, Unit 11.5 (2010).
 48. Afiahayati, Sato, K. & Sakakibara, Y. MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning. *DNA Res.* **22**, 69–77 (2015).

49. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014).
50. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* gr.186072.114- (2015). doi:10.1101/gr.186072.114
51. Bosi, E. *et al.* MeDuSa: a multi-draft based scaffolder. *Bioinformatics* **31**, 2443–51 (2015).
52. Tennessen, K. *et al.* ProDeGe: a computational protocol for fully automated decontamination of genomes. *ISME J.* **10**, 269–272 (2016).
53. Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
54. National Center for Biotechnology Information (US). *Extracting data from BLAST databases with blastdbcmd.* (2009).
55. He, S. *et al.* Metatranscriptomic array analysis of 'Candidatus Accumulibacter phosphatis'-enriched enhanced biological phosphorus removal sludge. *Environ. Microbiol.* **12**, 1205–17 (2010).
56. He, S. & McMahon, K. D. 'Candidatus Accumulibacter' gene expression in response to dynamic EBPR conditions. *ISME J.* **5**, 329–40 (2011).
57. Westra, E. R., Sünderhauf, D., Landsberger, M. & Buckling, A. Mechanisms and consequences of diversity-generating immune strategies. *Nat. Rev.*

- Immunol.* **17**, 719–728 (2017).
58. Weinberger, A. D. *et al.* Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* **8**, e1002475 (2012).
 59. van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).
 60. Paez-Espino, D. *et al.* CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio* **6**, e00262-15- (2015).
 61. Thompson, J. N. *The Coevolutionary Process*. **15**, (University of Chicago Press, 1994).
 62. Williams, H. T. *et al.* Phage-induced diversification improves host evolvability. *BMC Evol. Biol.* **13**, 17 (2013).
 63. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**, 43–51 (2011).
 64. Haq, I. U., Chaudhry, W. N., Akhtar, M. N., Andleeb, S. & Qadri, I. Bacteriophages and their implications on future biotechnology: a review. *Viol. J.* **9**, 9 (2012).
 65. Brockhurst, M. A., Koskella, B. & Zhang, Q.-G. in *Bacteriophages* 1–21 (Springer International Publishing, 2017). doi:10.1007/978-3-319-40598-8_7-1
 66. Lavelle, K. *et al.* A decade of *Streptococcus thermophilus* phage evolution in an Irish dairy plant. *Appl. Environ. Microbiol.* AEM.02855-17 (2018). doi:10.1128/AEM.02855-17
 67. Barr, J. J., Slater, F. R., Fukushima, T. & Bond, P. L. Evidence for

- bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS Microbiol. Ecol.* **74**, 631–42 (2010).
68. Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 1929–40 (2006).
 69. Bendall, M. L. *et al.* Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* **10**, 1–13 (2016).
 70. Zhou, Y., Pijuan, M., Zeng, R. J. & Yuan, Z. Involvement of the TCA cycle in the anaerobic metabolism of polyphosphate accumulating organisms (PAOs). *Water Res.* **43**, 1330–1340 (2009).
 71. OEHMEN, A. *et al.* Advances in enhanced biological phosphorus removal: From micro to macro scale. *Water Res.* **41**, 2271–2300 (2007).
 72. Wexler, M., Richardson, D. J. & Bond, P. L. Radiolabelled proteomics to determine differential functioning of *Accumulibacter* during the anaerobic and aerobic phases of a bioreactor operating for enhanced biological phosphorus removal. *Environ. Microbiol.* **11**, 3029–3044 (2009).
 73. Mino, T., van Loosdrecht, M. C. M. & Heijnen, J. J. Microbiology and biochemistry of the enhanced biological phosphate removal process. *Water Res.* **32**, 3193–3207 (1998).
 74. Welles, L. *et al.* Metabolic Response of “*Candidatus Accumulibacter Phosphatis*” Clade II C to Changes in Influent P/C Ratio. *Front. Microbiol.* **7**, 2121 (2017).

75. Crocetti, G. R. *et al.* Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantitation. *Appl. Environ. Microbiol.* **66**, 1175–82 (2000).
76. Carvalho, G., Lemos, P. C., Oehmen, A. & Reis, M. A. M. Denitrifying phosphorus removal: Linking the process performance with the microbial community structure. *Water Res.* **41**, 4383–4396 (2007).
77. Welles, L. *et al.* Accumulibacter clades Type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake. *Water Res.* **83**, 354–366 (2015).
78. Welles, L. *et al.* Metabolic Response of "Candidatus Accumulibacter Phosphatis" Clade II C to Changes in Influent P/C Ratio. *Front. Microbiol.* **7**, 2121 (2017).
79. Mao, Y., Yu, K., Xia, Y., Chao, Y. & Zhang, T. Genome Reconstruction and Gene Expression of "Candidatus Accumulibacter phosphatis" Clade IB Performing Biological Phosphorus Removal. *Environ. Sci. Technol.* **48**, 10363–10371 (2014).
80. Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. *ISME J.* **10**, 810–822 (2016).
81. Camejo, P. Y., Oyserman, B. O., McMahon, K. D. & Noguera, D. R. Integrated 'omic' analyses provide evidence that a Ca. Accumulibacter phosphatis strain performs denitrification under micro-aerobic conditions. *bioRxiv* 386516 (2018). doi:10.1101/386516

82. Joshi & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. (2011).
83. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–63 (2011).
84. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962 (2016).
85. Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* **40**, e155–e155 (2012).
86. Bosi, E. et al. MeDuSa: a multi-draft based scaffolder. *Bioinformatics* **31**, 2443–2451 (2015).
87. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
88. Warren, R. L. et al. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* **4**, 35 (2015).
89. Huntemann, M. et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.* **10**, 86 (2015).
90. Karp, P. D., Paley, S. & Romero, P. The Pathway Tools software. *Bioinformatics* **18 Suppl 1**, S225–32 (2002).
91. Caspi, R. et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34**, D511–6 (2006).
92. Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better

- detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324 (2008).
93. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
 94. Bushnell & Brian. BBMap: A Fast, Accurate, Splice-Aware Aligner. (2014).
 95. Zambelli, F. *et al.* RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res.* **46**, e46–e46 (2018).
 96. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
 97. Oyserman, B. O. O. *et al.* Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus Accumulibacter phosphatis*. *ISME J.* **10155**, 810–822 (2015).
 98. Welles, L. *et al.* Metabolic response of ‘*candidatus accumulibacter phosphatis*’ clade II C to changes in influent P/C ratio. *Front. Microbiol.* **7**, (2017).
 99. Marchler-Bauer, A. *et al.* CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
 100. Rehm, B. H. A. *Polyester synthases: natural catalysts for plastics.* *Biochem. J* **376**, (2003).

101. Shrivastav, A., Kim, H.-Y. & Kim, Y.-R. Advances in the applications of polyhydroxyalkanoate nanoparticles for novel drug delivery system. *Biomed Res. Int.* **2013**, 581684 (2013).
102. Martín, H. G. *et al.* Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.* **24**, 1263–1269 (2006).
103. He, S. & McMahon, K. D. 'Candidatus Accumulibacter' gene expression in response to dynamic EBPR conditions. *ISME J.* **5**, 329–340 (2010).
104. Comeau, Y., Hall, K. J., Hancock, R. E. W. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Res.* **20**, 1511–1521 (1986).
105. Hesselmann, R. P. X., von Rummell, R., Resnick, S. M., Hany, R. & Zehnder, A. J. B. Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Res.* **34**, 3487–3494 (2000).
106. Oehmen, A., Carvalho, G., Freitas, F. & Reis, M. A. Assessing the abundance and activity of denitrifying polyphosphate accumulating organisms through molecular and chemical techniques. *Water Sci. Technol.* **61**, 2061–2068 (2010).
107. Kim, J. M., Lee, H. J., Lee, D. S. & Jeon, C. O. Characterization of the denitrification-associated phosphorus uptake properties of 'Candidatus Accumulibacter phosphatis' clades in sludge subjected to enhanced biological phosphorus removal. *Appl. Environ. Microbiol.* **79**, 1969–79 (2013).

108. Wolfe, A. J. The acetate switch. *Microbiol. Mol. Biol. Rev.* **69**, 12–50 (2005).
109. Klein, A. H., Shulla, A., Reimann, S. A., Keating, D. H. & Wolfe, A. J. The intracellular concentration of acetyl phosphate in *Escherichia coli* is sufficient for direct phosphorylation of two-component response regulators. *J. Bacteriol.* **189**, 5574–81 (2007).
110. Miyake, M., Kataoka, K., Shirai, M. & Asada, Y. *Control of Poly-Hydroxybutyrate Synthase Mediated by Acetyl Phosphate in Cyanobacteria* Downloaded from. *JOURNAL OF BACTERIOLOGY* **179**, (1997).
111. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
112. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305 (2012).
113. Wilmes, P. *et al.* Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J.* **2**, 853–864 (2008).
114. Yagci, N. *et al.* Metabolic Model for Acetate Uptake by a Mixed Culture of Phosphate-and Glycogen-Accumulating Organisms Under Anaerobic Conditions. *Biotechnol Bioeng* **84**, 359–373 (2003).
115. Pramanik, J., Trelstad, P. L., Schuler, A. J., Jenkins, D. & Keasling, J. D. Development and validation of a flux-based stoichiometric model for enhanced biological phosphorus removal metabolism. *Water Res.* **33**, 462–476 (1999).

116. Silva, L. G. da *et al.* Revealing metabolic flexibility of *Candidatus Accumulibacter phosphatis* through redox cofactor analysis and metabolic network modeling. *bioRxiv* 458331 (2018). doi:10.1101/458331
117. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
118. Burow, L. C., Mabbett, A. N., McEwan, A. G., Bond, P. L. & Blackall, L. L. Bioenergetic models for acetate and phosphate transport in bacteria important in enhanced biological phosphorus removal. *Environ. Microbiol.* **0**, 070907134142001-??? (2007).
119. Saunders, A. M., Mabbett, A. N., McEwan, A. G. & Blackall, L. L. Proton motive force generation from stored polymers for the uptake of acetate under anaerobic conditions. *FEMS Microbiol. Lett.* **274**, 245–251 (2007).
120. Vemuri, G. N. & Aristidou, A. A. Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. *Microbiol. Mol. Biol. Rev.* **69**, 197–216 (2005).
121. Machado, D. *et al.* Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Comput. Biol.* **10**, e1003580 (2014).
122. Becker, S. A. *et al.* Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **4**, e1000082 (2008).
123. Henson, M. A. & Hanly, T. J. Dynamic flux balance analysis for synthetic microbial communities. *IET Syst. Biol.* **8**, 214–29 (2014).

