Title
Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis

Permalink
https://escholarship.org/uc/item/0qv3428b

Journal
ISME Journal, 10(4)

ISSN
1751-7362

Authors
Oyserman, BO
Noguera, DR
Del Rio, TG
et al.

Publication Date
2016-04-01

DOI
10.1038/ismej.2015.155

Peer reviewed
Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis

Ben O Oyserman1, Daniel R Noguera1, Tijana Glavina del Rio2, Susannah G Tringe2 and Katherine D McMahon1,3

1Department of Civil and Environmental Engineering, University of Wisconsin at Madison, Madison, WI, USA; 2US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA and 3Department of Bacteriology, University of Wisconsin at Madison, Madison, WI, USA

Previous studies on enhanced biological phosphorus removal (EBPR) have focused on reconstructing genomic blueprints for the model polyphosphate-accumulating organism Candidatus Accumulibacter phosphatis. Here, a time series metatranscriptome generated from enrichment cultures of Accumulibacter was used to gain insight into anaerobic/aerobic metabolism and regulatory mechanisms within an EBPR cycle. Co-expressed gene clusters were identified displaying ecologically relevant trends consistent with batch cycle phases. Transcripts displaying increased abundance during anaerobic acetate contact were functionally enriched in energy production and conversion, including upregulation of both cytoplasmic and membrane-bound hydrogenases demonstrating the importance of transcriptional regulation to manage energy and electron flux during anaerobic acetate contact. We hypothesized and demonstrated hydrogen production after anaerobic acetate contact, a previously unknown strategy for Accumulibacter to maintain redox balance. Genes involved in anaerobic glycine utilization were identified and phosphorus release after anaerobic glycine contact demonstrated, suggesting that Accumulibacter routes diverse carbon sources to acetyl-CoA formation via previously unrecognized pathways. A comparative genomics analysis of sequences upstream of co-expressed genes identified two statistically significant putative regulatory motifs. One palindromic motif was identified upstream of genes involved in PHA synthesis and acetate activation and is hypothesized to be a phaR binding site, hence representing a hypothetical PHA modulon. A second motif was identified ~35 base pairs (bp) upstream of a large and diverse array of genes and hence may represent a sigma factor binding site. This analysis provides a basis and framework for further investigations into Accumulibacter metabolism and the reconstruction of regulatory networks in uncultured organisms.

The ISME Journal (2016) 10, 810–822; doi:10.1038/ismej.2015.155; published online 10 November 2015

Introduction

Enhanced biological phosphorus removal (EBPR) is a widespread environmental biotechnology that exploits microorganisms capable of polyphosphate (polyP) accumulation to remove phosphorus (P) from wastewater (Bond et al., 1995; Hesselmann et al., 1999). The most widely studied organism responsible for EBPR in many wastewater treatment plants is named Candidatus Accumulibacter phosphatis (hereafter Accumulibacter; Nielsen et al., 2012). Although not yet isolated, a great deal has been learned about Accumulibacter physiology by studying enrichment cultures in laboratory scale bioreactors. Engineers have used this information to build quantitative metabolic models to predict how carbon, P, energy and reducing equivalents move through the wastewater ecosystem (Comeau et al., 1986; Oehmen et al., 2010). The accuracy and utility of these models depends heavily on an accurate understanding of Accumulibacter physiology.

It is well established that alternating cycles of carbon rich (feast) anaerobic and carbon poor (famine) aerobic environments are essential for successful EBPR operation (Oehmen et al., 2007). Under anaerobic conditions, Accumulibacter transports short-chain fatty acids (for example, acetate and propionate) into the cell and stores the carbon as polyhydroxyalkanoates (PHA). Current metabolic models generally assume that the energy for this process is obtained from ATP generated through polyP and glycogen degradation as well as from reducing equivalents generated through the
degradation of glycogen and the anerobic operation of the TCA cycle (Filipe and Daigger, 1998; Wexler et al., 2009; Zhou et al., 2009; Oehmen et al., 2010). Under subsequent aerobic conditions, PHA degradation supplies carbon and energy for growth and replenishment of glycogen and polyP storage molecules (Comeau et al., 1986; Mino et al., 1998).

The ability to store large quantities of polyP has led researchers to refer to organisms that display the aforementioned phenotype as polyphosphate-accumulating organisms. Although the ability to produce polyP and carbon storage polymers such as PHA and glycogen are phylogenetically dispersed traits (Wood and Clark, 1988; Reddy et al., 2003), by linking these metabolic processes and synchronizing them with key environmental conditions, Accumulibacter and other polyphosphate-accumulating organisms have a highly specialized and biotechnologically important phenotype that is the foundation of EBPR. To validate the underlying assumptions embedded in the metabolic models used by engineers, it is necessary to dissect the molecular mechanisms responsible for this synchronization.

The sequencing and completion of the first Accumulibacter genome (García Martín et al., 2006) has facilitated numerous transcriptional investigations, with the hypothesis that Accumulibacter’s highly coordinated physiology is the result of dynamic gene expression. Changes in transcript abundances have been previously investigated with reverse transcriptase quantitative PCR, microarrays, and RNA-seq under both stable and perturbed conditions (He et al., 2010; He and McMahon, 2011; Mao et al., 2014). However, these previous studies either targeted a handful of specific genes or examined limited time points during the anerobic–aerobic cycle. Here we used high-resolution time series metatranscriptomics with next-generation RNA-seq to identify highly expressed/dynamic genes and to identify putative co-regulated gene clusters. We then used comparative genomics to identify putative regulatory sequences and explore the underlying control mechanisms of the EBPR phenotype. Our results further validated some previously hypothesized aspects of Accumulibacter metabolism, uncovered important metabolic pathways that have been previously overlooked, and identified two putative sequence motifs providing the first step in determining gene expression regulatory mechanisms in Accumulibacter.

Materials and methods

Reactor maintenance

A single bioreactor was used in this study. Detailed reactor description and operating conditions are provided in García Martín et al. (2006). Briefly, the sequencing batch reactor was operated with a 2-l working volume and was fed with a mineral medium with acetate as a primary carbon source. The hydraulic retention time was 12 h and the sludge retention time was 4 days. The anerobic/aerobic cycle time was 6 h with 140 min anerobic contact (sparging with \( \text{N}_2 \) gas), 190 min aerobic contact (sparging with air) and 30 min settling time. Nitrification was inhibited using allylthiourea. For the experiment described herein, the cycle differed from normal operation in that acetate was fed over a 60-min period to elongate acetate contact. Representative phosphate, PHB and acetate profiles across the cycle are shown in Supplementary Figure 1A. Steady state operation is demonstrated by characteristic high anerobic and low aerobic P concentrations for the month before the experiment (Supplementary Figure 1B).

Chemical analysis

All chemical analyses were conducted during the same reactor cycle used for transcriptomic analysis (on 28 May 2013), except for hydrogen production assays, which were conducted after RNA-seq results were analyzed. To monitor the EBPR cycle, soluble phosphate, total suspended solids, volatile suspended solids and acetate were measured using previously described methods (Flowers et al., 2009). Polyhydroxyalkanoate analysis was performed using a gas chromatography–mass spectrometry as outlined previously (Comeau et al., 1988). Hydrogen production was measured under anerobic conditions using six batch tests conducted in 150-ml septum bottles with 25 ml of sludge. Three were negative controls in which no acetate was fed to get background hydrogen production rates and three were fed with 0.18 mmol of acetate. Hydrogen production was measured using a reduction gas analyzer (ta3000 Gas Analyzer, Trace Analytical, Ametek, Newark, DE, USA). To test the viability of glycine as a carbon source, anerobic batch tests were conducted in triplicate with negative control (no carbon addition), positive control (acetate addition) and glycine for a total of nine batch tests conducted in 60-ml septum bottles with 50 ml of sludge. Approximately 0.06 mmol of acetate and glycine were added respectively and phosphorus release was measured as previously described (Flowers et al., 2009).

Biomass sample collection and RNA extraction

Six biomass samples were collected across a single reactor cycle to capture key transition points in the EBPR cycle (Supplementary Figure 1A and Supplementary Table 1). Bulk biomass (2 ml) was collected in microcentrifuge tubes. Samples were centrifuged, supernatant removed and cell pellets flash frozen in dry ice and ethanol bath within 3 min of collection. RNA was extracted from the samples using an RNeasy kit (Qiagen, Valencia, CA, USA) with a DNase digestion step. RNA integrity and DNA contamination was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA).
Community characterization

Fluorescence in situ hybridization was conducted using PAOMIX probes to target all Accumulibacter clades (Crocetti et al., 2000), Acc-I-444 to target clade IA, and Acc-II-444 to target clade IIA, as previously described (Flowers et al., 2009). Cells were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). A Zeiss Imager.Z2 equipped with an AxioCam MRm camera was used to image fluorescing cells, which were then enumerated using ImageJ software (Abràmoff et al., 2004).

Library

Ribosomal RNA (rRNA) was removed from 1 μg of total RNA using Ribo-Zero rRNA Removal Kit (Bacteria) (Epicentre, Madison, WI, USA). Libraries were generated using the Truseq Stranded mRNA sample preparation kit (Illumina, San Diego, CA, USA). Briefly, the rRNA-depleted RNA was fragmented and reversed transcribed using Superscript II (Invitrogen, Carlsbad, CA, USA), followed by second-strand synthesis. The fragmented cDNA was treated with end-repair, A-tailing, adapter ligation and 10 cycles of PCR amplification.

Sequencing

The libraries were quantified using KAPA Biosystem’s next-generation sequencing library quantitative PCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq 2000 sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina’s cBot instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell was performed on the Illumina HiSeq 2000 sequencer using a TruSeq SBS sequencing kit 200 cycles, v3, following a 2 × 150 indexed run recipe. Sequence data were deposited at IMG/M under Taxon Object IDs 3300002341-3300002346.

Bioinformatics

Reads were quality trimmed and quality statistics were calculated using FASTX Toolkit (http://hannon lab.cshl.edu/fastx_toolkit/; Supplementary Figure 2). Ribosomal RNA sequences were removed with SortMeRNA using six built-in databases for bacterial, archaeal and eukaryotic small and large subunits (Kopylova et al., 2012). Reads that passed filtering were then mapped to the Accumulibacter clade IIA strain UW-1 (CAP2UW1; García Martín et al., 2006) chromosome and plasmids using the BWA mem algorithm with default parameters (Li and Durbin, 2009). Read counts were then calculated using HTseq with the ‘intersection strict’ parameter (Anders et al., 2014). Read counts were normalized by total reads in the sequencing run, the number of reads that remained after rRNA filtering, and the fraction of total reads that aligned to the Accumulibacter genome (Supplementary Table 1). Non-rRNA reads represented between 35 and 65 percent of all reads. Reads were then converted to log base two reads per kilo base per million (RPKM; Mortazavi et al., 2008; Supplementary Tables 4 and 5). Before clustering, all genes that did not have at least one observation with a log2(RPKM) read count of one or greater from the minimum observation were removed from the data set. Of 4735 genes in strain UW-1, 3893 passed this filter. These genes were further binned into clusters of co-expressed genes (Supplementary Table 6, Supplementary Figures S3–S7) based on an uncentered Pearson similarity metric of expression profiles followed by a centroid linkage clustering method using Java Treeview (Saldanha, 2004; Supplementary Figure 8). Clusters of co-expressed genes were then manually curated in Java TreeView. The resulting clusters are henceforth referred to as trend categories and are named with characters (for example, Trend Category A). Trend categories were then classified into patterns (Table 1) through visual inspection and comparison with known solute and biopolymer transformations that occur during an EBPR cycle.

Identification of highly expressed and highly dynamic genes

Genes that displayed the highest relative transcript abundance and those that displayed the largest changes in relative transcript abundance were identified as follows: each gene was represented as a vector of six RNA relative transcript abundance values. The maximum value of this vector represents

<table>
<thead>
<tr>
<th>Pattern</th>
<th>No. of genes</th>
<th>Trend categories</th>
<th>No. of trend categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anerobic acetate contact</td>
<td>126</td>
<td>Q,R</td>
<td>2</td>
</tr>
<tr>
<td>Redox transition</td>
<td>697</td>
<td>AA,BB,CC,DD,EE,FF,GG,HH,NN</td>
<td>9</td>
</tr>
<tr>
<td>Aerobic</td>
<td>1844</td>
<td>II,JK,KK,PP,QQ,RR,SS,TT,VV,WW,XX,YY,AAA,BBB,CCC,HHH,III,IIJ,</td>
<td>21</td>
</tr>
<tr>
<td>High phosphorus</td>
<td>40</td>
<td>O,P</td>
<td>2</td>
</tr>
<tr>
<td>Low phosphorus</td>
<td>438</td>
<td>F,EEE,GGG,NNN,OOO,PPP,QQQ,RRR,SSS,TTT,UUU,VVV</td>
<td>12</td>
</tr>
<tr>
<td>Sum</td>
<td>3145</td>
<td>—</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 1 A summary of the trend categories identified in this study and the patterns they display
the maximum relative expression of that gene. The maximum minus the minimum value of this vector represents the relative change in abundance over the entire cycle. Using these statistics, the genes may be ranked by those that show the highest relative expression and largest relative change. On the basis of the distribution of maximum expression values, a cut-off of 350 was determined to identify the highly expressed and highly dynamic genes (Supplementary Figure 9).

Functional enrichment analysis

Numerous subsets of genes were identified in this investigation including highly expressed/dynamic genes and trend categories. To determine whether these subsets were enriched in specific functions, a bootstrap method was used to determine how the distribution of COG functions compared with a null model produced from 1000 randomly generated gene subsets of equal size to the gene subset in question. The randomly generated null models for each gene subset was then compared with the observed abundance using a one sided t-test.

Operons and upstream motif identification

Putative operons were determined using the following set of criteria: (1) genes must have the same orientation; (2) adjacent genes were co-expressed with a cut-off correlation of 0.7; and (3) there was an intergenic region of 1000 base pairs or less (Supplementary Table 7). For each identified trend category, upstream sequences of called operons were analyzed for putative upstream motifs using MEME (Bailey et al., 2009; meme input.fasta -bfie background.txt –mod zoops –evt 0.05 –dna –nmotifs 10 –minsites 3 –o output), with one motif identified in a specific trend category (Supplementary Table 7). Additional motif sites were identified on the basis of sequence homology using MAST (Bailey et al., 1998; mast meme.txt input.fasta -bfie background.txt -oc. -nstatus -remcorr -ev 10 -norc -m 1). In addition, a motif search was conducted on the highly dynamic genes that displayed the aerobic acetate contact (AAC) pattern.

Results

Community composition, chemical analysis and total raw read statistics

On the date samples were collected for transcriptomics, P removal exceeded 99% and carbon and P dynamics characteristic of EBPR systems were observed (Supplementary Figure 1A). Accumulibacter relative abundance measured by fluorescence in situ hybridization was 80% of total DAPI-stained cells and Clade IIA accounted for 99% of the total Accumulibacter cells. P measurements at the end of aerobic and anerobic phases during the month of the investigation indicated stable state operation (Supplementary Figure 1B).

Illumina sequencing of ribosomal-depleted total RNA resulted in 1 461 769 869 reads across six samples (Supplementary Table 1). Quality filtering of reads removed 695 865 184, resulting in 765 904 685 for downstream analysis. Resulting reads were then mapped to the finished Accumulibacter clade IIA strain UW-1 reference genome (García Martín et al., 2006; Flowers et al., 2013), where 104 844 897 reads were aligned.

Co-expression patterns during a single EBPR cycle

Sampling across a single EBPR cycle and subsequent hierarchical clustering analysis allowed the identification of clusters of co-expressed transcripts henceforth referred to as trend categories. Trend categories had an average Pearson correlation of 0.96 and an average size of ~50 genes (Supplementary Table 8, Supplementary Figure 3). Ecologically relevant patterns of transcript abundance were identified that corresponded with the following important EBPR stages: high phosphorus concentration (Figure 1b), low phosphorus concentration (Figure 1c), AAC (Figure 1d), redox transition (Figure 1e), aerobic (Figure 1f). The largest number of transcripts displayed the aerobic pattern, followed by redox transition, low phosphorus, AAC and finally high-phosphorus patterns. A summary of the number of genes, trend categories and which trend categories were assigned to each pattern are given in Table 1. Ecologically relevant expression profile patterns were overlaid on a map of central Accumulibacter metabolic processes to enable interpretation of how they might relate to regulation of key pathways (Figure 2, see Supplementary Figure 11 for model with locus tags). Biochemical transformations were color-coded on the basis of the expression pattern to which the corresponding gene was assigned.

Key genes identified as upregulated after AAC include those involved in acetate activation, PHB synthesis (CAP2UW1_3191) and regulation (phasins, CAP2UW1_0642-CAP2UW1_0643), glycine cleavage (CAP2UW1_1955-CAP2UW1_1960), phospholipid monolayer formation (CAP2UW1_3192, CAP2UW1_3266, CAP2UW1_3702, CAP2UW1_0341, CAP2UW1_2586), carbonic anhydrase (CAP2UW1_1967) and hydrogenases (CAP2UW1_0998-CAP2UW1_0999, CAP2UW1_2286). The presence of high P was accompanied by relatively high expression rates of various transporters such as low-affinity P transporters (Pit, CAP2UW1_2085), sulfur transporters (SulP, CAP2UW1_2094) as well as porins (CAP2UW1_1151, CAP2UW1_1152) and the regulatory phoU (CAP2UW1_2086, CAP2UW1_2093, CAP2UW1_3728). Low P conditions corresponded to increased relative transcript abundance of Calvin cycle genes as well as those involved in high-affinity P transporters (Pst, CAP2UW1_2002-CAP2UW1_2008) numerous regulatory genes including phoR (CAP2UW1_1995), phoB (CAP2UW1_1996) and phoD (CAP2UW1_1732).
Differential transcript abundances across COG categories in a single EBPR cycle

Numerous subsets of genes (transcripts) were identified in this investigation including highly expressed/dynamic genes (Supplementary Table S10) as well as Trend Category Q and DD (Supplementary Table S6), in which genes related to Energy Production and Conversion were enriched ($P$-values $5.7e^{-26}$, $1.6e^{-10}$, $1.9e^{-13}$ and $2.3e^{-08}$, respectively) (Figures 3a and b). Furthermore, in each of these gene subsets, energy production and conversion represented the largest fraction of genes with predicted functions (Figures 3a and b). Additional details are located in the Supplementary Material.

Hydrogen gas production and glycine utilization in Accumulibacter

On the basis of the transcriptional profile of hydrogenases and a glycine cleavage operon detected in this study, we hypothesized that hydrogen gas production would occur during anerobic conditions after acetate addition and that glycine is a viable carbon source and would anerobic P release. To test these hypotheses, two sets of batch tests were conducted. Hydrogen gas production after acetate addition was measured and confirmed above background anerobic hydrogen gas production levels (Figure 4a). In addition, anerobic glycine addition resulted in P release, albeit at a lower rate than achieved by acetate contact (Figure 4b).
Figure 2 Updated metabolic model with biochemical reactions color-coded based on the expression profile pattern to which the corresponding gene was assigned. Genes involved in PHB formation demonstrate the anaerobic acetate contact pattern and are colored green. Genes involved in the TCA cycle/glycolysis generally demonstrated high expression levels across the redox transition (RT) and are colored blue. Genes involved in the Calvin Cycle demonstrated either the aerobic or low P patterns and are colored red and orange, respectively. Genes grouped into the high-phosphorus pattern are colored in yellow. These include low-affinity phosphate transporters. Ac, acetate; AcAc-CoA, acetoacetyl-CoA; Ac-CoA, acyl-CoA; Ac-AMP, acetyl AMP; Ac-P, acetyl-P; ADP-Glu, adenosine 5'-diphosphoglucose; CDPD, cytidine diphosphate diacylglycerol; C.I, complex I oxidative phosphorylation; C.II, complex II oxidative phosphorylation; C.III, complex III oxidative phosphorylation; C.IV, complex IV oxidative phosphorylation; E4-P, erythrose 4-phosphate; FNR, NADPH-ferredoxin reductase; Fru-1-6P, fructose 1,6-bisphosphate; Fru-6-P, fructose 6-phosphate; G3P, glyceraldehyde 3-phosphate; Glu, glucose; Glu-1-p, glucose 1-phosphate; Glu-6-P, glucose 6-phosphate; Gly, glycogen; GlyA, glycogen amylose; Glyc-P, glycerone-P; Long Chain FA, long chain fatty acid; PE, phosphatidylethanolamine; PGP, 1,2-diacyl-sn-glycerol-3p; pntAB, proton-translocating transhydrogenase; PolyP, polyphosphate; PPP, pyrophosphate-energized proton pump; Ptd-L-Ser, phosphatidylserine; Pyr, pyruvate; 1,3-bPG, 1,3-bisphosphoglyceric acid; Ri1,5P2, ribulose 1,5P2; Ri5-P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7-P, sedoheptulose-7-phosphate; SBP, sedoheptulose 1,7-bisphosphate; X5P, xylulose 5-phosphate; 3HB-CoA, (R)-3-hydroxy-butanoyl-CoA; 2-PG, 2-phosphoglycerate; 3-PG, 3-phosphoglyceric acid.

Figure Legend
- NADP
- NADPH
- NAD
- NADH
- Pi
- ADP
- ATP
- PPi
- Pi
- FDr
- AMP
- H2
- Redox Transition
- An Acetate Contact
- Aerobic
- High P
- Low P
- No Pattern

...... AAC Upstream Motif

Accumulibacter transcriptomics
BO Oyserman et al
815
The ISME Journal
Upstream sequence motif identification
To identify genes putatively co-regulated by cis-regulatory elements, an upstream motif analysis was conducted. A sequence motif was identified upstream of 51 sequences within Trend Category DD using MEME (each with a \( P \)-value \( <9.07 \times 10^{-4} \)) and upstream of 25 additional genes using MAST (each with a \( P \)-value \( <1.44 \times 10^{-4} \); Figure 5a, Supplementary Table S9; Bailey et al., 2009, 1998). A majority of the sequence motifs (~60%) were found between 25 and 45 base pairs upstream of the start codon (Figure 5b). In addition, a motif analysis using MEME conducted on the subset of genes within the AAC that were designated as highly dynamic revealed a palindromic motif upstream of 10 genes (each with a \( P \)-value \( <1.58 \times 10^{-6} \)) and an additional 5 genes were identified with MAST (each with a \( P \)-value \( <1.20 \times 10^{-4} \); Figure 5c, Supplementary Table S9; Bailey et al., 2009, 1998). The genes found to share this upstream motif were generally (~50%) between 45 and 95 base pairs downstream; however, the spacing ranged considerably (Figure 5d).

Discussion
Carbon metabolism in Accumulibacter
One of the defining features of Accumulibacter metabolism is anerobic intracellular carbon flux. Bulk analysis of Accumulibacter-enriched cultures...
consistently demonstrates that as VFAs are transported into the cell, PHAs are synthesized and glycogen is degraded (Oehman et al., 2007). Therefore, we expected to find genes involved in the flux of carbon through acetyl-CoA and then to PHB to be upregulated during AAC. Interestingly, no evidence for an immediate transcriptome-level response upregulating genes involved in acetate acquisition via active (acetate permease, acIP, CAP2UW1_1608) or passive (porins, CAP2UW1_1151 and CAP2UW1_1152) transport was identified upon initial acetate contact (that is, early in the cycle). However, acetate uptake triggered the upregulation of numerous other genes directly related to intracellular acetate and PHA processing. Once inside the cell, acetate is activated via the low-affinity acetate phosphotransferase (CAP2UW1_1002) and high-affinity acyl-coenzyme A (CoA) synthetase (CAP2UW1_3266) to acetyl-CoA, both of these transcripts exhibit the AAC pattern (Figures 1d and 2). Other pathways to acetyl-CoA also exhibited the AAC pattern including pyruvate kinase (CAP2UW1_1298) and phosphoenolpyruvate carboxykinase (CAP2UW1_1298) and an anerobic glycine cleavage system operon (CAP2UW1_1955-CAP2UW1_1960; Figures 1d and 2). These findings suggest that Accumulibacter transcriptionally regulates gene expression to route as much carbon as possible toward acetyl-CoA formation via several different pathways. This is discordant with conventional metabolic models for Accumulibacter that aim to identify a single (or primary) route.

For example, the anerobic expression of the glycine cleavage system suggests an important role for glycine, and potentially other non-VFAs carbon sources, in Accumulibacter metabolism. Some previously reported experimental evidence from full scale systems provides support for this hypothesis: glycine addition resulted in the highest P release of any tested amino acid in batch tests with activated sludge (Wiliński, 2009). Furthermore, we conducted batch tests using Accumulibacter-enriched sludge and confirmed that glycine addition results in P release (Figure 4b). Thus, although free glycine contributes to the carbon budget of Accumulibacter, it remains unclear whether Accumulibacter may be able to liberate glycine from more complex sources. For example, one possible complex source of glycine would be collagen, which has an important role in bacterial biofilm formation (Oliver-Kozup et al., 2013, 2011). Every third amino acid in bacterial collagen is composed of glycine (Yu et al., 2014). Interestingly, a predicted collagenase/peptidase operon (CAP2UW1_0989-CAP2UW1_0991) was highly expressed and clustered within the AAC. Together the glycine cleavage and collagenase/peptidase operon represent a possible mechanism for the release and acquisition of glycine via peptidase/collagenase activity and subsequent anerobic glycine oxidation, thus providing additional acetate and reducing equivalents through Strickland reactions (Sagers and Gunsalus, 1960; Okamura-Ikeda et al., 1993; Andreesen, 1994). The specificity of this predicted collagenase/peptidase should be further investigated to understand its role during anerobic carbon metabolism in Accumulibacter. No mechanism for collagen synthesis was identified in Accumulibacter; however, collagen and other peptides produced by either Accumulibacter or the other bacterial community members may provide an important and previously unrecognized carbon source anerobically.

Regardless of its initial form, once the carbon source has been transformed into acetyl-CoA, three additional enzymes are required for PHA synthesis: β-ketothiolase (PhaA), acetoacetyl-CoA reductase.
specialized metabolism of Accumulibacter, and during anerobic carbon metabolism, energy generation and conversion reactions must provide the reducing equivalents and ATP needed to drive the specialized metabolism of Accumulibacter, and much research effort has been directed at quantifying these (Comeau et al., 1986; Mino et al., 1998; Oehmen et al., 2007). Indeed, a salient characteristic of the highly expressed and highly dynamic gene subsets as well as Trend Category Q and DD was the enrichment of genes found in the Energy Production and Conversion COG Category (Figures 3a and b). We were particularly intrigued by the fact that multiple oxidoreductases were among the most highly expressed and most dynamic genes assigned to these patterns, since oxidoreductases are key to managing the movement of reducing equivalents through the cell. This prompted us to look more closely at their potential role in Accumulibacter metabolism as revealed through expression patterns.

Current EBPR models explicitly require that anerobic reducing equivalents for PHA synthesis be produced through glycogen degradation (during oxidation of glyceraldehyde-3-phosphate) and/or anerobic operation of the TCA cycle (oxidation of isocitrate, alpha-ketoglutarate and malate), in the form of NADH (Hesselmann et al., 2000; Zhou et al., 2009). When intracellular stores of glycogen are limiting, acetate flux through the TCA cycle increases via the oxidation of isocitrate, alpha-ketoglutarate and malate (Zhou et al., 2009). However, reducing equivalents provided by glycolysis and the TCA cycle are in the form of NADH while PHA synthesis generally requires NADPH (Peoples and Sinskey, 1989; Steinbüchel, 2009). How- ever, reducing equivalents provided by glycolysis and the TCA cycle are in the form of NADH while PHA synthesis generally requires NADPH (Peoples and Sinskey, 1989; Steinbüchel, 2009). This redundancy is accomplished by hydrogenase activity that produces free H₂, (2) NADPH is the cofactor required for PhaB activity, and transhydro- genase activity is involved in the conversion of NADH to NADPH, (3) the supply of NADH is greater than what is required or cannot be converted to NADPH quickly enough and (4) in addition to requiring reducing equivalents, other factors such as anerobic demand for glycolysis-derived ATP (Saunders et al., 2007) drive glycogen degradation.

Therefore, we hypothesize that upon glycogen degradation (1) the NADH/NAD⁺ ratio increases and recycling of NAD⁺ is accomplished by hydrogenase activity that produces free H₂, (2) NADPH is the cofactor required for PhaB activity, and transhydro- genase activity is involved in the conversion of NADH to NADPH, (3) the supply of NADH is greater than what is required or cannot be converted to NADPH quickly enough and (4) in addition to requiring reducing equivalents, other factors such as anerobic demand for glycolysis-derived ATP (Saunders et al., 2007) drive glycogen degradation.

In support of (2), we confirmed that NAD(P)H transhydrogenase (CAP2UW1_4179, CAP2UW1_4180) was highly expressed. However, curiously the two
subunits show slight relative upregulation in the aerobic phase. Enzymatic assays testing the NADH/NDPH specificity of Accumulibacter PhaB and measurements of NAD(P)H/NAD(P) ratios throughout an EBPR cycle should also help test these hypotheses.

**Low/high soluble phosphorus and carbon correlated gene expression**

During aerobic metabolism within an EBPR cycle both nutrient rich (feast) and poor (famine) states exist. Immediately after first oxygen contact, the environment is nutrient rich; P and C are abundant extracellularly and intracellularly (as stored PHB) respectively. As the aerobic phase continues, P is transported and stored intracellularly while PHB is degraded to drive P uptake, polyP/glycogen synthesis and growth. Thus, at the end of the aerobic phase, both P and C may be considered limiting as extracellular P and PHB are depleted.

The high-phosphorus pattern (Figure 1b) included the most highly transcribed gene in Accumulibacter, a porin operon (CAP2UW1_1151, CAP2UW1_1152), as well as inorganic phosphate transporters (Pit, CAP2UW1_2085) and sulfur transporters within two operons (CAP2UW1_2085-CAP2UW1_2087, CAP2UW1_2092-2094) (Figure 2). The high relative abundance of Pit, SuIP and porin transcripts during periods of high soluble P concentrations is consistent with the high flux of metabolites, such as P and counter cations, into and out of the cell across the anerobic/aerobic phases of an EBPR cycle. The Pit/ SuIP operons also contained multiple phoU-like genes (Liu et al., 2005; Oganesyan et al., 2005). Co-expression of PhoU and Pit during periods with high soluble P concentrations are consistent with PhoU’s hypothesized role as a negative regulator for the PhoR-PhoB two-component system (Baek et al., 2007). During periods with low P, phoU transcripts decreased and the transcription of the high-affinity P transporter system occurred (Pst, CAP2UW1_2002-CAP2UW1_2008) as well as phoR (CAP2UW1_1995), phoB (CAP2UW1_1996) and phoD (CAP2UW1_1732) (Figure 2).

When P levels are low at the end of the aerobic phase, measured PHB levels are also below detection (Supplementary Figure 1). During this low-phosphorus (Figure 1C) or ‘carbon starved’ state, we identified an increase in the relative abundance of Ribulose 1,5-bisphosphate carboxylase (Rubisco; CAP2UW1_0825). Multiple genes involved in the Calvin cycle exhibited a similar expression pattern (CAP2UW1_0959, CAP2UW1_0958, CAP2UW1_0957, CAP2UW1_0823) (Figure 2). The potential for carbon fixation by the Accumulibacter lineage has been hypothesized since the original metagenome was sequenced (García Martín et al., 2006) and further supported by the sequencing of additional draft genomes (Skennerton et al., 2014). In addition, experiments have shown that Accumulibacter enrichments may be sustained in the absence of organic substrates in the medium (Kang and Noguera, 2014). Investigations into the ability of Accumulibacter to fix carbon during aerobic conditions when PHAs have been exhausted must be conducted to confirm this potential. Furthermore, perturbation experiments decoupling low P and low C conditions will be important in further distinguishing co-expression patterns.

**New insights into regulatory mechanisms in accumulibacter**

**Upstream sequence motif identification.** Dynamic gene expression is an adaptive mechanism that allows organisms to respond to changing environmental conditions (Seshasayee et al., 2009). By regulating genes at the level of transcription, organisms may prevent the energy waste associated with the synthesis of unnecessary proteins (Stoebel et al., 2008), or the negative effects of certain proteins when synthesized under the inappropriate conditions (Eames and Kortemme, 2012). Co-expression clusters determined through transcriptomic analysis may be used in conjunction with comparative genomics methods to identify sequence motifs that represent putative regulatory features (Kellis et al., 2003; Imam et al., 2014). We hypothesized that genes following a particular expression profile would be under the same regulatory mechanisms and may therefore share upstream sequence motifs.

To explore such regulatory features, we analyzed the upstream sequences of operons within each trend category. A sequence motif was identified upstream of 51 sequences within Trend Category DD and a sequence homology search identified an additional 25 locations of the motif. (Figure 5a, Supplementary Table S9). We hypothesize that this motif is a putative sigma factor binding site because a large number of the motif occurrences are found ~35 bp upstream (Figure 5b) of the transcriptional start site (Harley and Reynolds, 1987). No strong -10 site was identified in this study, suggesting that the promoter associated with this putative DNA binding domain may have a reduced requirement for a -10 binding region (Thouverot et al., 2004; Hook-Barnard and Hinton, 2007). However, the biological relevance of this motif must be tested using additional methods such as DNase-seq (He et al., 2014).

We also identified a palindromic sequence motif upstream of 10 genes binned in the AAC pattern and subsequent homology searches identified an additional six locations of the motif (Figure 5c). Genes found to have this upstream motif included those involved directly in PHA synthesis, such as phaC (CAP2UW1_3191), phaP (CAP2UW1_0642-CAP2UW1_0643), phaR (CAP2UW1_3918), both low and high-affinity acetate activation enzymes (CAP2UW1_1002, CAP2UW1_3266), a potassium transporter (CAP2UW1_2100), and an operon containing the cytoplasmic hydrogenase (CAP2UW1_
production and conversion (Figures 3a and b) display circular functions, such as those related to energy levels of expression. Further, we showed that partitioned times during the cycle rather than sustained mRNA productivity at specific, environmentally triggered times throughout the cycle are maintained via bursts of high expression. Transcriptomic analysis of Accumulibacter under diverse conditions will result in further differentiating clusters of co-expressed genes, giving additional power to comparative methods to identify putative regulatory mechanisms.

The identification of the first putative regulatory motifs in Accumulibacter represents a milestone that will invite further investigation to (1) determine whether the sequence motifs are true transcription factor binding sites, (2) identify the transcriptional regulatory protein associated with these sites, (3) determine whether regulation is via activation or repression. Transcriptomic analysis of Accumulibacter under diverse conditions will result in further differentiating clusters of co-expressed genes, giving additional power to comparative methods to identify putative regulatory mechanisms.

Previous proteomics-based research has demonstrated that relative protein abundances do not change markedly across an EBPR cycle (Wilmes et al., 2008, 2008; Wexler et al., 2009). In contrast, we demonstrated that the stable protein abundances observed through the cycle are maintained via bursts of high mRNA productivity at specific, environmentally triggered times during the cycle rather than sustained levels of expression. Further, we showed that particular functions, such as those related to energy production and conversion (Figures 3a and b) display highly dynamic transcripts over an EBPR cycle.

### Conclusion

Metatranscriptomic sequencing was conducted on six samples collected across an aerobic/anerobic EBPR cycle in a bioreactor enriched in Accumulibacter clade IIA. Despite relatively stable protein abundance identified in previous studies (Wexler et al., 2009), the identification of numerous co-expressed gene sets provides strong evidence that transcriptional regulation is critical for the anaerobic/aerobic metabolism of Accumulibacter. AAC triggered the expression of genes related to acetyl-CoA and PHB formation, including genes involved with aerobic glycine metabolism. These findings suggest that Accumulibacter routes more diverse carbon sources through acetyl-CoA to PHB than previously recognized.

The discovery of hydrogenase expression and the demonstration of hydrogen gas production highlight previously unknown components of the EBPR cycle, and indicate that a redox imbalance may exist during AAC. We suggest that reducing equivalents in the form of NADH from glycogen degradation cannot be used directly for PHB synthesis, but rather NADPH is required by PhaB for the reduction of acetoacetyl-CoA to (R)-3-hydroxybutyryl-CoA, as commonly observed in other organisms. Thus, efficient conversion of NADH to NADPH may be a rate-limiting step that has not been adequately recognized in metabolic models of Accumulibacter (Niederholtmeyer et al., 2010; Angermayr et al., 2012). This is an excellent example of how discovery-based sequencing can reveal new metabolic features.

Comparative genomics of upstream sequences in co-regulated gene sets allowed, for the first time, the identification of two sequence motifs in Accumulibacter. The first was a palindromic motif upstream of genes showing upregulation during AAC and involved in acetate activation, PHA granule formation, fatty acid synthesis, counter cation transport and reducing equivalent balance through hydrogen gas production. The second motif was identified as a putative sigma factor binding site upstream of many genes/operons within the redox transition pattern. The discovery of this putative regulatory motif suggests that portions of acetate uptake, PHA synthesis, and P metabolism (counter cation transport) are co-regulated. Additional metatranscriptomic analyses may further identify regulatory mechanisms within Accumulibacter and the regulons associated with the unique Accumulibacter metabolism.

### Conflict of Interest

The authors declare no conflict of interest.

### Acknowledgements

We thank the following individuals for insightful discussion and friendly reviews of draft manuscripts: Christopher Lawson, Francisco Moya and Travis Korosh. We thank Alisha Truman, Mitch Heffernan, Antonio Garcia and Lianne Estrella for assistance with bioreactor operation. KDM acknowledges funding from the US National Science Foundation (CBET-0967646) and the UW-Madison Graduate School. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### References


Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)