Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics

Elizabeth A. McDaniel[1*†], Francisco Moya-Flores[2,3,4*†], Natalie Keene Beach[2,6], Pamela Y. Camejo[2,5], Ben O. Oyserman[7,8], Matthew Kizaric[2], Eng Hoe Khor[2], Daniel R. Noguera[2,9], and Katherine D. McMahon[1,2†]

[1] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI, USA

[3] Geroscience Center for Brain Health and Metabolism (GERO), Santiago, Chile

[4] Millenium Initiative for Collaborative Research on Bacterial Resistance (MICROB-R), Iniciativa Científica Milenio, Chile

[5] Millennium Institute for Integrative Biology (iBio), Santiago, Chile

[6] Carollo Engineers, Inc., Broomfield, CO, USA

[7] Bioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands

[8] Microbial Ecology, Netherlands Institute of Ecological Research, Wageningen, The Netherlands

[9] DOE Great Lakes Bioenergy Research Center, Madison, WI, USA

* equal contribution

† Corresponding Authors: elizabethmcd93@gmail.com, fmoya@wisc.edu, and kdmcmahon@wisc.edu

ABSTRACT

Natural microbial communities consist of closely related taxa that may exhibit phenotypic differences and inhabit distinct niches. However, connecting genetic diversity to ecological properties remains a challenge in microbial ecology due to the lack of pure cultures across the microbial tree of life. *'Candidatus* Accumulibacter phosphatis' is a polyphosphate-accumulating organism that contributes to the Enhanced Biological Phosphorus Removal (EBPR) biotechnological process for removing excess phosphorus from wastewater and preventing eutrophication from downstream receiving waters. Distinct Accumulibacter clades often co-exist in full-scale wastewater treatment plants and lab-scale enrichment bioreactors, and have been hypothesized to inhabit distinct ecological niches. However, since individual strains of the Accumulibacter lineage have not been isolated in pure culture to date, these predictions have been made solely on genome-based comparisons and enrichments with varying strain composition. Here, we used genome-resolved metagenomics and metatranscriptomics to explore the activity of co-existing Accumulibacter strains in an engineered bioreactor environment. We obtained four high-quality genomes of Accumulibacter strains that were present in the bioreactor ecosystem, one of which is a completely contiguous draft genome scaffolded with long Nanopore reads. We identified core and accessory genes to investigate how gene expression patterns differed among the dominating strains. Using this approach, we were able to identify putative pathways and functions that may confer distinct functions to Accumulibacter strains and provide key functional insights into this biotechnologically significant microbial lineage.

47 **IMPORTANCE**

48 '*Candidatus* Accumulibacter phosphatis' is a model polyphosphate accumulating
49 organism that has been studied using genome-resolved metagenomics,
50 metatranscriptomics, and metaproteomics to understand the EBPR process. Within the
51 Accumulibacter lineage, several similar but diverging clades are defined by shared
52 sequence identity of the polyphosphate kinase (*ppk1)* locus. These clades are predicted
53 to have key functional differences in acetate uptake rates, phage defense mechanisms,
54 and nitrogen cycling capabilities. However, such hypotheses have largely been made
55 based on gene-content comparisons of sequenced Accumulibacter genomes, some of
56 which were obtained from different systems. Here, we performed time-series genome-
57 resolved metatranscriptomics to explore gene expression patterns of co-existing
58 Accumulibacter clades in the same bioreactor ecosystem. Our work provides an approach
59 for elucidating ecologically relevant functions based on gene expression patterns
60 between closely related microbial populations.
61
62
63
64
65
66
67
68
69

INTRODUCTION

Naturally occurring microbial assemblages are composed of closely related species or subpopulations and harbor extensive genetic diversity. Coherent lineages are often comprised of distinct clades, ecotypes, or “backbone subpopulations” defined by high within-group sequence similarities that translate to ecologically relevant diversity (1–3). For example, genetically diverse ecotypes of the marine cyanobacterium *Prochlorococcus* that are 97% identical by their 16S ribosomal RNA gene sequences exhibit markedly different light-dependent physiologies and distinct seasonal and geographical patterns (1, 2, 4, 5). Closely related *Nitrospira* representing species-like lineages that coexisted for long periods of time in a full-scale wastewater treatment plant exhibited functional differences in preferences for nitrite concentrations and carbon substrates (6). However, dissecting the emergent ecological properties of species-like coherent microbial lineages is still a grand challenge in microbial ecology given that these principles are likely not universal among the breadth of microbial diversity, and is further hindered when there is a lack of pure cultures (7–9).

Enhanced Biological Phosphorus Removal (EBPR) is an economically and environmentally significant process for removing excess phosphorus from wastewater. This process depends on polyphosphate accumulating organisms (PAOs) that polymerize inorganic phosphate into intracellular polyphosphate. ‘*Candidatus* Accumulibacter phosphatis’ (hereafter referred to as Accumulibacter) is a member of the *Betaproteobacteria* in the *Rhodocyclaceae* family and has long been a model PAO (10, 11). Accumulibacter achieves net phosphorus removal through cyclical anaerobic-aerobic phases of feast and famine. In the initial anaerobic phase, abundantly available volatile

93 fatty acids (VFAs) such as acetate are converted to polyhydroxyalkanoates (PHAs), a
94 carbon storage polymer. However, PHA formation requires significant ATP input and
95 reducing power, which is supplied through polyphosphate and glycogen degradation. In
96 the subsequent aerobic phase in which oxygen is available as a terminal electron
97 acceptor and soluble carbon sources are depleted, Accumulibacter uses PHA reserves
98 for growth. Accumulibacter recovers energy during aerobic PHA degradation by
99 replenishing polyphosphate and glycogen reserves, which can be later used for future
100 PHA formation as the anaerobic/aerobic cycle repeats. This feast-famine oscillation
101 through sequential anaerobic-aerobic cycles operating on minute-to-hourly time scales
102 gives rise to net phosphorus removal and the overall EBPR process (10, 12). A few
103 'omics-based studies have revealed potential regulatory modules that could coordinate
104 Accumulibacter's complex physiological behavior during these very dynamic
105 environmental conditions (13–16), but few have attempted to assign unique gene
106 expression dynamics to specific clades (17).

107 Accumulibacter can be subdivided into two main types (I and II) and further
108 subdivided into multiple clades based upon polyphosphate kinase (*ppk1*) sequence
109 identity since the 16S rRNA marker is too highly conserved among the breadth of known
110 Accumulibacter lineages to resolve species-level differentiation (18–21). The two main
111 types share approximately 85% nucleotide identity across the *ppk1* locus (21) which
112 mirrors genome-wide average nucleotide identity (ANI) (22). The persistence of several
113 phylogenetically distinct clades within the Accumulibacter lineage in both full-scale
114 wastewater treatment plants and lab-scale enrichment bioreactors suggests that different
115 clades may inhabit distinct ecological niches (17, 23). This is largely supported by

116 comparative genomics of Accumulibacter metagenome-assembled genomes (MAGs)
117 recovered from different bioreactor systems and classified using the *ppk1* locus (14, 17,
118 22, 24–27). Members of different clades have been hypothesized to vary in their acetate
119 uptake rates, nitrate reduction capacity, and phage defense mechanisms (22, 26). For
120 example, clade IA seems to exhibit higher acetate uptake rates and phosphate release
121 rates, whereas evidence exists that clade IIA can reduce nitrate but clade IA cannot (25,
122 28). However, the full breadth of ecological differentiation among Accumulibacter clades
123 is not yet clear.

124 Here, we used genome-resolved metagenomics and time-series
125 metatranscriptomics to investigate gene expression patterns of co-existing
126 Accumulibacter strains (representing two main clades) during a standard EBPR cycle.
127 We assembled four high-quality genomes of the dominant (clades IA and IIC) and low
128 abundant (clades IIA and IIF) Accumulibacter strains present in the bioreactor ecosystem.
129 Additionally, we constructed a contiguous draft genome of an Accumulibacter clade IIC
130 genome achieved using long Oxford Nanopore Technology (ONT) reads, representing
131 one of the highest-quality genomes for this clade recovered to date. Using these
132 assembled genomes, we performed time-series RNA-sequencing throughout a typical
133 EBPR feast-famine cycle to study expression profiles of different Accumulibacter strains,
134 with an emphasis on core and flexible gene content. More generally, our work reveals
135 putative ecological functions of co-existing taxa within a dynamic ecosystem.

136

137

138

139 **MATERIALS AND METHODS**

140 **Bioreactor Operation and Sample Collection**

141 A laboratory-scale sequence batch reactor (SBR) was seeded with activated
142 sludge from the Nine Springs Wastewater Treatment Plant (Madison WI, USA) during
143 August 2015. The bioreactor was operated to simulate EBPR as described previously
144 (24). Briefly, an SBR with a 2-L working volume was operated with the established
145 biphasic feast-famine conditions: a 6-h cycle consisted of the anaerobic phase (sparging
146 with N_2 gas), 110-min; aerobic (sparging with oxygen), 180-min; settle, 30-min; draw and
147 feed with an acetate-containing synthetic wastewater feed solution, 40-min. A reactor
148 hydraulic residence time of 12-h was maintained by withdrawing 50% of the reactor
149 contents after the settling phase, then filling the reactor with fresh nutrient feed; a mean
150 solids retention time of 4-d was maintained by withdrawing 25% of the mixed reactor
151 contents each day immediately prior to a settling phase. Allylthiourea was added to the
152 synthetic feed solution to inhibit ammonia oxidation and indirectly inhibit
153 nitrification/denitrification.

154 Three biomass samples for metagenomic sequencing were collected in
155 September of 2015 by centrifuging 2-mL of mixed liquor at 8,000 *x g* for 2 min, and DNA
156 was extracted using a phenol:chloroform extraction protocol. Seven samples for RNA
157 sequencing were collected throughout a single EBPR cycle (3 in the anaerobic phase, 4
158 in the aerobic phase) in October of 2016 following the sampling strategy conducted by
159 Oyserman et al. (13). Samples were flash-frozen on liquid nitrogen immediately after
160 centrifuging and discarding the supernatant. RNA was extracted using a TRIzol-based
161 extraction method (Thermo Fisher Scientific, Waltham, MA) followed by phenol-

162 chloroform separation and RNA precipitation. RNA was purified following an on-column
163 DNAse digestion using the RNase-Free DNase Set (Qiagen, Venlo, Netherlands) and
164 cleaned up with the RNeasy Mini Kit (Qiagen, Venlo, Netherlands). Accumulibacter clade
165 quantification was performed on biomass samples collected on the same day of the
166 metatranscriptomics experiment using clade-specific *ppk1* qPCR primers as described by
167 Camejo et al. (23).

168 **Metagenomic and Metatranscriptomic Library Construction and Sequencing**

169 Metagenomic libraries were prepared by shearing 100 ng of DNA to 550 bp long
170 products using the Covaris LE220 (Covaris) and size selected with SPRI beads (Beckman
171 Coulter). The fragments were ligated with end repair, A-tailing, Illumina compatible
172 adapters (IDT Inc) using the KAPA-Illumina library preparation kit (KAPA Biosystems).
173 Libraries were quantified using the KAPA Biosystem next-generation sequencing library
174 qPCR kit and ran on a Roche LightCycler 480 real-time PCR instrument. The quantified
175 libraries were prepared for sequencing on the Illumina HiSeq platform using the v4
176 TruSeq paired-end cluster kit and the Illumina cBot instrument to create a clustered flow
177 cell for sequencing. Shotgun sequencing was performed at the University of Wisconsin –
178 Madison Biotechnology Center DNA Sequencing facility on the Illumina HiSeq 2500
179 platform with the TruSeq SBS sequencing kits, followed by 2x150 indexing. Raw
180 metagenomic data consisted of 219.4 million 300-bp Illumina HiSeq reads with
181 approximately 3.9 Gbp per sample. Nanopore sequencing was performed according to
182 the user's manual version SQK-LSK108.

183 Total RNA submitted to the University of Wisconsin-Madison Biotechnology Center
184 was verified for purity and integrity using a NanoDrop 2000 Spectrophotometer and an

185 Agilent 2100 BioAnalyzer, respectively. RNA-Seq paired-end libraries were prepared
186 using the TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA). Each sample was
187 processed for ribosomal depletion, using the Ribo-Zero™ rRNA Removal Kit (Bacteria).
188 mRNA was purified from total RNA using paramagnetic beads (Agencourt RNAClean XP
189 Beads, Beckman Coulter Inc., Brea, CA) and fragmented by heating in the presence of a
190 divalent cation. The fragmented RNA was then converted to cDNA with reverse
191 transcriptase using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad,
192 California, USA) with random hexamer priming and the resultant double-stranded cDNA
193 was purified. cDNA ends were repaired, adenylated at the 3' ends, and then ligated to
194 Illumina adapter sequences. The quality and quantity of the DNA were assessed using
195 an Agilent DNA 1000 series chip assay and Thermo Fisher Qubit dsDNA HS Assay Kit.
196 Libraries were diluted to 2nM, pooled in an equimolar ratio, and sequenced on an Illumina
197 HiSeq 2500, using a single lane of paired-end, 100 bp sequencing, and v2 Rapid SBS
198 chemistry.

199 **Metagenomic Assembly and Annotation**

200 We applied two different assembly and binning approaches to obtain high-quality
201 genomes of the dominant Accumulibacter strains present in the reactors. For the first
202 approach, Illumina unmerged reads were quality filtered and trimmed using the Sickle
203 software v1.33 (29). Reads were merged with FLASH v1.0.3 22 (30), with a mismatch
204 value of ≤0.25 and a minimum of ten overlapping bases from paired sequences, resulting
205 in merged read lengths of 150 to 290 bp. FASTQ files were then converted to FASTA
206 format using the Seqtk software v1.0 (31). Metagenomic reads from all three samples
207 were then co-assembled using the Velvet assembler with a k-mer size of 65 bp, a

208 minimum contig length of 200 bp, and a paired-end insert size of 300 bp (32). Metavelvet
209 was used to improve the assembly generated by Velvet (33). Metagenomic contigs were
210 then binned using Maxbin (34). This approach allowed us to assemble high-quality
211 genomes of clades IIA and IIC, named UW5 and UW6, respectively. Long Nanopore
212 reads were generated under the standard protocol and used to scaffold assembled
213 contigs with MeDuSa (35), and manually inspected to remove short contigs and
214 decontaminated using ProDeGe (36) and Anvi'o (37). Further scaffolding was performed
215 on both of these genomes using Nanopore long reads using LINKS (38) and GapCloser
216 (39).

217 Because this pipeline did not produce high-quality genomes of clades IA and IIF,
218 we applied a different approach to assemble these genomes. All reads from each
219 metagenomic sample were individually and co-assembled assembled using metaSPAdes
220 (40). Reads from each metagenome were mapped against each of the assemblies using
221 bbmap with a 95% sequence identity cutoff to obtain differential coverage (41), and
222 contigs were binned using MetaBat (42). Identical clusters of bins were dereplicated
223 across assemblies using dRep (43) to obtain the best quality genome of both clades IA
224 and IIF, named UW4 and UW7, respectively. All genomes were quality checked using
225 CheckM (44) and functional annotations assigned with Prokka and KofamKOALA (45,
226 46). We were unable to scaffold the genomes of UW4-IA and UW7-IIF with long Nanopore
227 reads because of their low abundance at the time of metagenomic sequencing.

228 Each genome was assigned to a particular clade both by comparing the *ppk1*
229 sequence identity and ANI to previously published Accumulibacter genomes. A *ppk1*
230 database was created using sequences from previously published Accumulibacter

231 references and clone sequences. We searched for the *ppk1* gene in our draft genomes
232 using this database and aligned the corresponding hits with MAFFT (47). A phylogenetic
233 tree of aligned *ppk1* gene sequences from Accumulibacter references, select clone
234 sequences, and the outgroups *Dechloromonas aromatica RCB* and *Rhodoc*yclus *tenuis*
235 *DSM 110* with RAxML v8.1 with 100 rapid bootstraps (48). To use any given
236 Accumulibacter reference genome in downstream comparisons, a confident *ppk1* hit had
237 to be identified using the above methods. As a technical note, the IA-UW2 strain
238 assembled by Flowers et al. (22) was renamed to UW3, as the original assembly
239 contained an additional contig that was likely from a prophage. The prophage contig was
240 removed from the IA-UW2 strain and renamed to UW3, and therefore the UW numerical
241 nomenclature follows in sequential order after UW3. A phylogenetic tree of *ppk1*
242 nucleotide sequences was performed and compared to select clone sequences. Pairwise
243 genome-wide ANI was calculated between the four Accumulibacter draft genomes and
244 all available Accumulibacter genomes in Genbank with FastANI (49). A species tree of all
245 Accumulibacter references and outgroup genomes was constructed using single-copy
246 markers from the GTDBK-tk (50), aligned with MAFFT (47), built with RAxML with 100
247 rapid bootstraps (48), and visualized in iTOL (51). To compare genome-wide ANI scores
248 of Accumulibacter references with pairwise nucleotide identity of the *ppk1* locus of all
249 references, we performed pairwise BLAST for all *ppk1* coding regions and reported the
250 percent identity (52).

251 **Identification of Core and Accessory Gene Content**

252 We identified core and accessory gene content of Accumulibacter clades through
253 clustering orthologous groups of genes (COGs) using PyParanoid (53). We used the four

254 Accumulibacter MAGs generated in this study (Table 1) and high-quality Accumulibacter
255 reference genomes (above 90% completeness and less than 5% redundancy) for this
256 analysis, with *Dechloromonas aromatica RCB* and *Rhodocyclus tenuis DSM 110* as
257 outgroups (Accumulibacter references listed in Supplementary Table 1 available at
258 https://figshare.com/articles/dataset/Accumulibacter_References_Metadata/13237148).
259 Briefly, an all-vs-all comparison of proteins for all genomes was performed using
260 DIAMOND (54), pairwise homology scores calculated with the InParanoid algorithm (55),
261 gene families constructed with MCL (56), and HMMs built for each gene family with
262 HMMER (57). A phylogenetic tree was constructed for genomes used for the COGs
263 analysis by using the coding regions of the *ppk1* locus and overlaid with the
264 presence/absence of core and flexible gene content using the ete3 python package (58).
265 A similarity score between UW Accumulibacter genomes was calculated as the pairwise
266 Pearson correlation coefficient between gene families using the numpy python package
267 (59). The presence and absence of different gene families were visualized in an upset
268 plot between UW Accumulibacter genomes using the ComplexUpset package based on
269 the UpSetR package in R (60). All data wrangling and plotting were performed using the
270 tidyverse suite of packages in R (61).

271 **Metatranscriptomics Mapping and Processing**

272 Metatranscriptomics reads from each of the seven samples were quality filtered
273 using fastp (62) and rRNA removed with SortMeRNA (63). The coding regions of all four
274 Accumulibacter genomes assembled in this study (Table 1) were predicted with Prodigal
275 (64) and annotated with both Prokka and KofamKOALA (45, 46), and concatenated
276 together to create a mapping index. Metatranscriptomic reads were competitively

277 pseudoaligned to the reference index and counts quantified with kallisto (65). Genes with
278 a sum of more than 10,000 counts across all 7 samples were removed, as these are likely
279 ribosomal RNAs that were not removed previously. Additionally, genes annotated as
280 rRNAs by barrnap were manually removed (45).

281 To explore genes and metabolic functions that may exhibit hallmark anaerobic-
282 aerobic feast-famine cycling patterns among the dominant IA-UW4 and IIC-UW6 strains,
283 we identified sets of genes that were differentially expressed between anaerobic and
284 aerobic conditions using DESeq2 (66). Differential gene expression calculations were
285 performed for each strain separately to account for differences in the relative abundance
286 of metatranscriptomic reads mapping back to each of the four MAGs. Low-count genes
287 were removed individually for each genome by requiring that all genes had to have more
288 than 10 counts mapped across 3 or more samples. For IA-UW4, 3549 genes of a total
289 3919 predicted genes remained after applying this filter, and for IIC-UW6, 4297 genes of
290 a total 4877 predicted genes remained after applying this filter. We analyzed expression
291 patterns of differentially expressed genes using a threshold cutoff of ± 1.5 log2 fold-
292 change between anaerobic and aerobic conditions and then identified differentially
293 expressed genes as core or accessory among clades IA (str. UW4) and IIC (str. UW6).
294 For the core group analysis, the corresponding ortholog in each genome had to meet the
295 minimum count threshold as described above to include in the comparison, with genes in
296 one or both of the strains being differentially expressed. For the accessory analysis for
297 each strain, the gene had to be differentially expressed and not contained in the other
298 genome but could be present in other strains or other genomes belonging to that clade.

299 **Data and Code Availability**

300 Raw sequencing files for the three metagenomes and seven metatranscriptomes
301 and genome assemblies for the four Accumulibacter genomes are available at NCBI
302 under the BioProject accession PRJNA668760. All genome assemblies used in this study
303 in their current forms as well as transcriptomic count tables are available at:
304 https://figshare.com/projects/Metabolic_Plasticity_of_Accumulibacter_Clades/90614. All
305 code is available at https://github.com/elizabethmcd/R3R4.
306

307 **RESULTS**

308 **Enrichment of Co-Existing Accumulibacter Clades**

309 We inoculated a 2-L sequencing-batch reactor with activated sludge from a full-
310 scale wastewater treatment plant in Madison, WI, USA and maintained operation for 40
311 months. The reactor was primarily fed with acetate and operated under cyclic anaerobic-
312 aerobic phases with a 4-day solids retention time to simulate the EBPR process and
313 enrich for Accumulibacter (Figure 1). During a period of stable EBPR (14 months following
314 inoculation) in which acetate was completely diminished by the start of the aerobic phase
315 and soluble phosphate was below 1.0 mg-L^{-1} at the end of the aerobic phase, we collected
316 seven samples across a single cycle (three in the anaerobic phase, four in the aerobic
317 phase) for RNA-sequencing (Figure 1A). We performed qPCR of the *ppk1* as described
318 previously (23) to quantify Accumulibacter clades present in the reactor on the same day
319 as the RNA-seq experiment (Figure 1B). At the time of the RNA-seq experiment, the
320 bioreactor was primarily enriched in Accumulibacter clade IA, followed by clade IIC and
321 clade IIA.

322 From metagenome samples collected from this bioreactor earlier, we assembled
323 high-quality genomes of belonging to four different Accumulibacter clades (Table 1),
324 including the dominant IA and IIC (strains IA-UW4 and IIC-UW6, respectively) clades and
325 the less abundant IIA and IIF clades (strains IIA-UW5 and IIF-UW7, respectively) (Figure
326 2). Each of the assembled genomes falls within the established Accumulibacter clade
327 nomenclature as defined by *ppk1* sequence identity and phylogenetic placement when
328 compared to other publicly available Accumulibacter genomes and clone sequences
329 (Figure 2A). Additionally, the *ppk1* hierarchical structure is also reflected by pairwise-
330 average nucleotide identity (ANI) boundaries between clades, with some exceptions for
331 newly assembled Accumulibacter genomes that fall outside of the established '*C.*
332 Accumulibacter phosphatis' lineage (Figure 2B). All assembled genomes are above 90%
333 complete and contain less than 5% redundancy as calculated by CheckM (44) (Table 1).
334 We were able to scaffold the assemblies of strains IIA-UW5 and IIC-UW6 using long
335 Nanopore reads, constructing a completely contiguous draft genome of the IIC-UW6
336 strain. To our knowledge, the assembled IIC-UW6 genome is the most contiguous,
337 highest-quality reference genome available for this clade, providing a valuable new
338 Accumulibacter reference genome (Figure 2C).

339 **Distribution of Shared and Flexible Gene Content between Clades**

340 We characterized the functional diversity of different Accumulibacter clades
341 through clustering orthologous groups of genes (COGs) (Figure 3A). We repeated a
342 COGs analysis from Skennerton et al. (26) and Oyserman et al. (67), but included an
343 updated set of high-quality Accumulibacter references, the UW-generated genomes, and
344 the outgroups *Dechloromonas aromatica* and *Rhodocyclus tenuis* (Figure 3). The

345 Accumulibacter lineage exhibits a tremendous amount of functional diversity, with only
346 approximately 25% of COGs shared amongst all Accumulibacter clades (Figure 3A).
347 Additionally, there is extensive diversity within the two Accumulibacter types and within
348 individual clades, with only a few representatives of each clade more than approximately
349 75% similar to each other by gene content (Supplementary Figure 1). Specifically, clade
350 IIC broadly harbors the most diversity of all sampled genomes by gene content, with
351 representatives of this clade not sharing more than 25% of COGs (Figure 3A). Other
352 clades are more similar by gene content, but this could be due to fewer genomes sampled
353 from these clades. Most of the highest-quality, publicly available Accumulibacter genome
354 references belong to clades IIF and IIC.

355 We then compared COGs between Accumulibacter clades collected from
356 bioreactors seeded from the Nine Springs Wastewater Treatment Plant in Madison, as
357 these genomes represent a shared origin (Figure 3B). In addition to two lab-scale
358 bioreactors harboring distinct Accumulibacter clades (strains UW1 and UW3, and UW4-
359 7, respectively) (22, 24), we recently characterized clade IC (strain UW-LDO-IC) using
360 genome-resolved metagenomics and metatranscriptomics (14). Overall, genomes in
361 Type I and Type II cluster respectively, where genomes within Type I are overall more
362 similar to each other, as expected (Supplementary Figure 1). Clade IIA genomes (strains
363 UW1 and UW5) were assembled from separate enrichments derived from the same full-
364 scale treatment plant approximately 12 years apart, but are very similar by COG
365 presence/absence (Supplementary Figure 1). UW genomes within Type II are less similar
366 by gene content and seem to be more diverse in this regard, though again this could be
367 due to under-sampling of Type I (8 genomes in Type I and 14 in Type II). Additionally,

368 clade IIF contains more overlap in gene content with Type I genomes in gene families
369 than some Type II genomes do to each other (Supplementary Figure 1).
370 COG overlap and uniqueness is shown by the intersection of COG groups between
371 UW Accumulibacter clades and the *Rhodocyclus* and *Decholoromonas* outgroups (Figure
372 3B). There are 1075 gene families shared among all UW Accumulibacter clades and
373 outgroup genomes, with 277 separate gene families only within the UW Accumulibacter
374 clades (Figure 3B). Among individual clades, there are 887 groups only within clade IIC
375 (UW6), 399 within clade IIA (UW1 and UW5), 399 within IIF (UW7), 292 within clade IA
376 (UW3 and UW4), and 196 within IC (UW-LDO). Although clade IIC str. UW6 contains
377 more overlap in gene content with SK-02 (Figure 3A), the intersection analysis (Figure
378 3B) identified gene families that are available to individual strains within the same
379 bioreactor ecosystem, highlighting the potential for functional differentiation.
380 **Transcriptional Profiles of Accumulibacter Strains**
381 Although *ppk1* locus quantification showed that the bioreactor was mostly enriched
382 in clade IA (see Figure 1B), surprisingly, more RNA-seq reads mapped to the clade IIC-
383 UW6 genome than to the clade IA-UW4 (Figure 4A). Approximately 40 million total
384 transcriptional reads aligned to the strain IIC-UW6 genome across the full time-series
385 experiment, whereas strains IA-UW4 and IIA-UW5 recruited roughly the same number of
386 reads, with 14 million and 11 million, respectively (Table 2). Markedly lower read levels
387 mapped to the IIF-UW7 strain relative to the other three clades, and therefore we do not
388 include the IIF-UW7 strain in our subsequent analyses. We explored the top differentially
389 expressed genes in strain IIC-UW6, comparing anaerobic and aerobic conditions (Figure
390 4B). Genes upregulated in the anaerobic phase belonged to pathways for central-carbon

391 transformations, fatty acid biosynthesis, and PHA synthesis. Interestingly, all genes within
392 the high-affinity phosphate transport system Pst were upregulated in the aerobic phase.
393 This includes the substrate-binding component *pstS, phoU* regulator, and the ABC-type
394 transporter *pstABC* complex (Figure 4B).

395 The co-existing clades in the bioreactor system differ in both the presence of sets
396 of nitrogen cycling gene sets and their expression profiles (Figure 4C). In strain IA-UW4,
397 we detected the *napAGH* genes for nitrate reduction, whereas we did not detect *napB*,
398 possibly due to these genes being close to the end of a contig in this genome. We also
399 detected a *nirS*-like nitrite reductase*,* both subunits of the nitric oxide reductase *norQD,*
400 and the nitrous oxide reductase *nosZ* in strain IA-UW4. Most of these genes in strain IA-
401 UW4 exhibit the highest expression patterns in the early and late anaerobic stages
402 (Figure 4C). Conversely, strain IIC-UW6 contains the *narGHIJ* machinery for nitrate
403 reduction, a different *nirS*-like nitrite reductase than in IA-UW4, and does not contain a
404 nitrous oxide reductase. However, strain IIC-UW6 does contain the nitrite reductase *nirBD*
405 for reducing nitrite to ammonium, which is uniformly expressed across the cycle (Figure
406 4C).

407 **Expression Dynamics of Genes Shared by Strains IA-UW4 and IIC-UW6**

408 When characterizing expression dynamics of COGs shared between strains IA-
409 UW4 and IIC-UW6, we found very few orthologs that exceeded the 1.5-fold change cutoff
410 for differential expression in both strains (Figure 5). These genes include a
411 coproporphyrinogen III oxidase, NAD-reducing HoxS subunits, the *phaC* subunit of the
412 PHA synthase, the long-chain fatty-acid-CoA ligase FadD13, and several hypothetical
413 proteins. The fatty-acid CoA-ligase incorporates ATP, CoA, and fatty acids of variable

414 length to form acyl-CoA to be degraded for energy production, incorporated into complex
415 lipids, or used in other metabolic pathways (68). The acyl-CoA ligase is upregulated in
416 the anaerobic phase of both clades, suggesting a role in anaerobic carbon metabolism,
417 as acyl-CoA can ultimately form acetyl-CoA through beta-oxidation. The *phaC* subunit of
418 the PHA synthase is differentially expressed in the anaerobic phase for both clades, as
419 is typical of the hallmark EBPR metabolism for Accumulibacter (13, 69).

420 Additionally, both the IA-UW4 and IIC-UW6 strains carry most of the subunits for
421 the bidirectional NiFe Hox system, encoding an anaerobic hydrogenase. The HoxH and
422 HoxY beta and delta subunits form the hydrogenase moiety, and the FeS-containing
423 HoxF and HoxU alpha and gamma subunits catalyze NAD(P)H/NAD(P)+
424 oxidation/reduction coupled to the hydrogenase moiety (70–72). In both clades, the FeS
425 cluster HoxF and HoxU alpha and gamma subunits are strongly differentially expressed
426 in the anaerobic phases, as was observed previously in a genome-resolved
427 metatranscriptomic investigation of strain IIA-UW1 (13). However, the IIC-UW6 genome
428 is missing the HoxY delta subunit of the hydrogenase moiety, and only strain IA-UW4
429 exhibits differential expression of the HoxH beta subunit in the anaerobic phase. Strain
430 IIC-UW6 does exhibit higher expression of HoxH in the late anaerobic and early aerobic
431 phases but does not meet the threshold cutoff for differential gene expression. A prior
432 study focused on strain IIA-UW1 also demonstrated hydrogenase activity in the anaerobic
433 phase, and then demonstrated hydrogen gas production after acetate addition (13). The
434 hydrogen gas production was hypothesized to replenish NAD+ after glycogen
435 degradation. Subsequent comparative genomics analysis suggested that the Hox system
436 is an ancestral state of Accumulibacter (67). Furthermore, recent metabolic models which

437 integrated hydrogen production under periods of excess acetate or glycogen were better
438 able to capture the diversity of observed EBPR stoichiometry (73). In this ecosystem,
439 both strains may exhibit anaerobic hydrogenase activity and thus produce hydrogen, but
440 due to the missing subunit in the IIC-UW6 genome and lack of differential gene
441 expression, this is uncertain.

442 Interestingly, only one strain exhibited differential expression patterns of several
443 core ortholog subsets important for EBPR-related functions exhibit. For example, subunits
444 of the high-affinity phosphate transporter Pst system are only differentially expressed in
445 strain IIC-UW6 and not IA-UW4, as observed for strain IIC-UW6 above (see Figure 4B).
446 Additionally, nitrogen cycling genes that contain orthologs in both clades display different
447 expression profiles. A nitrite reductase is only differentially expressed in the anaerobic
448 phase in strain IA-UW4 and the denitrification protein NirQ is differentially expressed in
449 only strain IIC-UW6. Whereas the nitrogen fixation protein subunit NifU is differentially
450 expressed in both strains. To activate acetate to acetyl-CoA, the high-affinity acetyl-CoA
451 synthase (ACS) or low-affinity acetate kinase/phosphotransacetylase (AckA/Pta)
452 pathways can be employed. The strain IA-UW4 assembly is missing an acetate kinase,
453 and phosphate acetyltransferase is only differentially expressed in this strain but not in
454 IIC-UW6. Within strain IIC-UW6, both the acetate kinase and phosphate acetyltransferase
455 genes exhibit relatively low levels of expression over the time-series. Both strains contain
456 several acetyl-CoA synthases that are highly expressed in the anaerobic phase,
457 particularly upon acetate contact, but do not exhibit notable differences in differential
458 expression between the two phases. Previous work based on community proteomics and
459 enzymatic assays have suggested that although both activation pathways are present

460 and expressed in Accumulibacter, the high-affinity pathway is preferentially used (74, 75).
461 Although a phosphate acetyltransferase is differentially expressed in strain IA-UW4, this
462 gene has lower transcripts relative to the acetyl-CoA synthase, providing more evidence
463 that the lineage (specifically both clades IA and IIC) may employ the high-affinity system.

464 **Differential Expression of Flexible Genes in Strain IA-UW4**

465 We lastly characterized expression dynamics of flexible gene content in strain IA-
466 UW4 (Figure 6). Since a majority of the differentially expressed flexible genes in strain
467 IIC-UW6 were annotated as hypothetical proteins, we focused on putative differentiating
468 features of clade IA-UW4. Flexible gene content for IA-UW4 can be defined as any
469 ortholog that is not present in the IIC-UW6 genome, whether or not that ortholog is only
470 contained within clade IA genomes, Type I as a whole, within other Type II or clade IIC
471 genomes, or the *Dechloromonas* or *Rhodocyclus* outgroups. Thus, the absence of a
472 particular ortholog in the strain IIC-UW6 genome but present in the IA-UW4 genome or
473 any other genomes was inferred as a lack of functional capability of IIC-UW6 in this
474 particular ecosystem.

475 Particularly striking is the differential expression profiles of nitrogen cycling genes
476 that do not contain orthologs in the IIC-UW6 genome. As stated above, clade IA-UW4
477 contains the *nap* system whereas strain IIC-UW6 contains the *nar* system for reducing
478 nitrate to nitrite (Figure 4C). In strain IA-UW4, the *napDGH* subunits are differentially
479 expressed between the anaerobic and aerobic phases, and these genes are present in
480 the outgroups, as well as both Accumulibacter types except the IIC-UW6 genome (Figure
481 6). Both genomes contain a *nirS* nitrite reductase, but the respective homologs are not
482 within the same COG, and thus strain IA-UW4 contains a different *nirS* that is upregulated

483 in the anaerobic phase. Strain IIC-UW6 does not contain both subunits of the nitric-oxide
484 reductase *norQD,* whereas strain IA-UW4 does. These subunits are not differentially
485 expressed above the set threshold but do exhibit the highest expression upon acetate
486 contact in the anaerobic phase (see Figure 4B). Additionally, strain IIC-UW6 does not
487 contain the nitrous-oxide reductase *nosZ* (see Figure 4B), whereas strain IA-UW4 does,
488 and this is differentially expressed in the anaerobic phase (Figure 6). These results
489 suggest that strain IA-UW4 may have the capability for full denitrification from nitrate to
490 nitrogen gas, or at least use the reduction of nitrate/nitrite as a source of energy in the
491 anaerobic phase in a denitrifying bioreactor. Although strain IIC-UW6 does contain the
492 *nirBD* nitrite reductase complex for reducing nitrite to ammonium, these genes are
493 uniformly expressed throughout the cycle (see Figure 4B).
494

495 **DISCUSSION**

496 In this work, we performed time-series transcriptomics to explore gene expression
497 profiles of Accumulibacter strains belonging to clades IA and IIC, which co-existed within
498 the same bioreactor ecosystem. By applying genome-resolved metagenomic and
499 metatranscriptomic techniques, we were able to postulate niche-differentiating features
500 of distinct Accumulibacter clades based on core and accessory gene content and
501 differential expression profiles. We assembled high-quality genomes from four
502 Accumulibacter clades and used long reads to generate a single contiguous scaffold of
503 the IIC-UW6 genome. By performing an updated clustering of orthologous groups of
504 genes analysis, we were able to better demonstrate the similarity and uniqueness among
505 Accumulibacter strains from this system. Unexpectedly, although clade IA dominated in

506 abundance by *ppk1* locus quantification, clade IIC was more transcriptionally active. By
507 comparing the expression profiles of COGs shared between strains IA-UW4 and IIC-
508 UW6, we found that genes involved in EBPR-related pathways may differ between the
509 strains with respect to transcription-level regulation. Notably, strain IIC-UW6 showed a
510 strong upregulation of the high-affinity Pst phosphorus transport system in the aerobic
511 phase. High transcription of the Pst system in IIC-UW6 towards the end of the aerobic
512 phase when phosphorus concentrations were lower suggests that the IIC-UW6 strain was
513 more sensitive to bulk phosphate concentrations. This would be congruent with
514 specialization to relatively low phosphorus conditions. Whereas lack of differential
515 expression of the Pst system in IA-UW4 suggests this strain is more specialized under
516 high phosphorus conditions. Strain IA-UW4 contains most of the subunits for the
517 individual key steps for denitrification, and exhibits strong anaerobic upregulation of these
518 genes, suggesting a differentiating role for strain IA-UW4 in nitrogen cycling under
519 denitrifying conditions. Additionally, the IA-UW4 strain contains all subunits for encoding
520 the Hox anaerobic hydrogenase system, which exhibits upregulated gene expression in
521 the anaerobic phase. However, the IIC-UW6 strain is missing one of the subunits even
522 though the other subunits were expressed (but not differentially). Production of hydrogen
523 gas could be a differentiating role for the IA-UW4 strain, as this would impact potential
524 interactions with diverse flanking community members in addition to replenishing NAD+
525 (13, 24, 76).

526 The Accumulibacter lineage harbors extensive diversity between and amongst
527 individual clades which is exhibited not only by *ppk1*-based phylogenies, but also by
528 pairwise genome-wide ANI and COG comparisons. For example, strain IIA-UW1 and

529 strain IA-UW3 were present in the same bioreactor ecosystem and only share 85% ANI
530 (22). In the bioreactor enrichment from this study, the dominant strains IA-UW4 and IIC-
531 UW6 share less than 80% ANI (Figure 2B). Recently, a cutoff of >95% ANI has been
532 suggested to delineate distinct bacterial species or cohesive genetic units (49), which is
533 based on observations of coverage gaps at 90-95% sequence identity when
534 metagenomic reads were mapped back to assembled MAGs (77–80). These sequence
535 cutoffs have also been benchmarked against homologous recombination rates and ratios
536 of nonsynonymous to synonymous (*dN/dS*) nucleotide differences (81). Among
537 Accumulibacter genomes adhering to the established *ppk1*-based phylogeny, genomes
538 within individual clades fall within the >95% ANI threshold, whereas ANI comparisons
539 within and between Types I and II are more variable (Figure 2B). Genomes within Type I
540 are similar by approximately 90% ANI, whereas genomes within Type II are not as
541 coherent.

542 The apparent coexistence of diverse Accumulibacter clades within an ecosystem
543 then raises questions of ecological roles, putative interactions, and evolutionary dynamics
544 of this lineage over time and space. The ecotype model of speciation suggests that
545 cohesive, ecologically distinct ecotypes can coexist because they occupy separate niche
546 spaces (3, 82). Because Accumulibacter has not been isolated in pure culture to date,
547 most ecological and evolutionary inferences have been made from genome-based
548 comparisons (26, 67) or mapping metagenomic reads back to assembled genomes with
549 minimum percent identity cutoffs that may not sufficiently distinguish among sub-clades
550 (83, 84). However, the apparent diversity within the Accumulibacter lineage may require
551 reevaluation of the phylogeny and taxonomy overall. Although the *ppk1*-based clade

552 designations coincide with genome-wide ANI based cutoffs remarkably well
553 (Supplementary Figure 2B), there are some discrepancies. A few Accumulibacter
554 genomes have been assembled that do not adhere to the *ppk1*-based population
555 structure but fall within the Accumulibacter lineage based on a species tree of single-copy
556 core genes (Supplementary Figure 2A), such as the proposed novel species '*Candidatus*
557 Accumulibacter aalborgensis' (85). Additionally, a few genomes preliminarily classified by
558 the Genome Taxonomy Database (GTDB) as Accumulibacter branch outside of the
559 established nomenclature and are quite divergent based on the species and *ppk1*
560 phylogenies as well as genome-wide and *ppk1* sequence similarities (Figure 2B,
561 Supplementary Figure 2). Overall, substantial work is needed to reevaluate the
562 phylogenetic structure of the Accumulibacter lineage as a whole and to connect with
563 signatures of homologous recombination and evolutionary trajectories within coexisting
564 clades to understand the genetic boundaries and thus ecological interdependencies of
565 individual clades or strains.

566 Although the boundaries of individual Accumulibacter clades or strains have not
567 been entirely resolved to infer separate ecotypes, denitrification capabilities of specific
568 clades have been repeatedly suggested to be a niche-differentiating feature (14, 23, 25).
569 Previously, two separate bioreactors fed with either acetate or propionate as the sole
570 carbon source were enriched with different Accumulibacter morphotypes, which were also
571 linked to different denitrification capabilities (28). A study with a bioreactor highly enriched
572 in clade IC inferred simultaneous use of oxygen, nitrite, and nitrate as electron acceptors
573 under micro-aerobic conditions either due to metabolic flexibility or heterogeneity within
574 IC members (23). The resident strain IC-UW-LDO also possessed the full suite of genes

575 for denitrification and expressed them in microaerobic conditions (14). Coexisting
576 Accumulibacter strains with differing denitrifying abilities have also been observed in a
577 denitrifying EBPR bioreactor enriched in clade IA and clade IC (17, 27). The clade IA
578 strain encoded genes for complete denitrification, whereas the co-occurring clade IC
579 strain and other flanking community members did not (27). Additionally, Gao et al.
580 performed an ancestral genome reconstruction analysis similar to Oyserman et al. 2017
581 (67) to specifically understand the flux of denitrification gene families (27). The *nir* gene
582 cluster was identified as a core COG to all sampled Accumulibacter genomes, whereas
583 the periplasmic nitrate reductase *napAGH* and nitrous oxide reductase *nosZDFL* clusters
584 were core to only Type I Accumulibacter genomes, and inferred to have been lost for
585 some Type II genomes (27). Additionally, Wang et al. subsequently observed that an
586 internal stop codon within the open reading frame of the *nosZ* gene in a clade IC strain
587 coincided with decreased transcripts mapping back (17, 27). In our study, we observed
588 differential expression of denitrification genes in strain IA-UW4 but not strain IIC-UW6.
589 These expression differences could explain why the two strains can coexist and
590 contribute to overall ecosystem functioning (e.g. phosphorus removal) by using different
591 electron acceptors in the anaerobic/aerobic conditions. Although the synthetic feed for
592 the bioreactors in this study inhibits ammonia oxidation through allylthiourea addition (and
593 thus inhibits nitrite/nitrate production to fuel respiratory denitrification), the expression
594 profiles of denitrification genes in strain IA-UW4 suggests this strain is equipped to have
595 a differentiating role in a denitrifying bioreactor. We specifically observed differential
596 expression of nitrate reductase and nitrous-oxide reductase genes in strain IA-UW4,
597 where homologs of these genes were not present in the strain IIC-UW6 genome and other

598 Type II genomes. From our work and previous studies, lack of the complete genetic
599 repertoire for denitrification among and within individual clades in addition to
600 transcriptional-based evidence suggests that denitrification capabilities could be a niche
601 differentiating role for specific Accumulibacter populations.

602 Accumulibacter's anaerobic metabolism has also been shown to differ under
603 varying substrate or stoichiometric conditions and possibly between the two
604 Accumulibacter types. Recent simulations of 'omics datasets paired with enzymatic
605 assays hint that the routes of anaerobic metabolism for Accumulibacter may differ based
606 on environmental conditions (73). Under high acetate conditions, polyphosphate
607 accumulation arises through the glyoxylate shunt, whereas under increased glycogen
608 concentrations, glycolysis is used in conjunction with the reductive branch of the TCA
609 cycle (73). Furthermore, Welles et al. found that Type II Accumulibacter can switch to
610 partial glycogen degradation more quickly than Type I, enabling Type II to fuel VFA uptake
611 in polyphosphate-limited conditions (86, 87). Conversely, when concentrations of
612 polyphosphate are high, VFA uptake rates are higher in Type I. However, these studies
613 only identified Accumulibacter lineages based on *ppk1* types, and therefore extrapolated
614 metabolic assumptions from presumably more resolved strains to the entire type. These
615 broad interpretations may obscure apparent functional differences between the two types
616 and individual clades. Future experiments could integrate multi-'omics approaches to
617 investigate not only metabolic flexibility under different substrate conditions and
618 perturbations but also how metabolisms are distributed amongst Accumulibacter clades
619 in bioreactor ecosystems that are not completely enriched in a single clade or strain and
620 elucidate key functional differences.

621 **CONCLUSIONS**

622 In this study, we integrated genome-resolved metagenomics and time-series
623 metatranscriptomics to understand gene expression patterns of co-existing
624 Accumulibacter strains within a bioreactor ecosystem. We found evidence for
625 denitrification gene expression in the dominating IA-UW4 strain, but higher overall
626 transcriptional activity for the IIC-UW6 strain along with differential expression of a high-
627 affinity phosphorus transporter. The coexistence of Accumulibacter populations exhibiting
628 different gene expression patterns for phosphorus and nitrogen cycling and hydrogen
629 productions suggest niche differentiating features and contribution to different metabolic
630 programs. Detailed experiments incorporating microautoradiography combined with
631 fluorescence in situ hybridization (MAR-FISH) could be applied to confirm metabolic
632 activities of coexisting Accumulibacter populations and how they contribute to important
633 EBPR-related functions (88). These results also highlight the need to reevaluate the
634 phylogenetic structure of the Accumulibacter lineage and define how potential genetic
635 boundaries may translate to individual ecological niches that allow for the coexistence of
636 multiple clades, species-like groups, or strains. Additionally, we highlight an approach for
637 exploring functions of co-existing species-like uncultivated microbial lineages by exploring
638 gene expression patterns of shared and accessory genes. Overall, this work emphasizes
639 the need to understand the basic ecology and evolution of microorganisms underpinning
640 biotechnological processes, which can lead to better treatment process control and
641 resource recovery outcomes in the future (89).

642

643

ACKNOWLEDGEMENTS

We thank Dr. Jeffrey Lewis for feedback on the differential expression analyses and interpretation. This work was supported by funding from the National Science Foundation (MCB-1518130) to K.D.M. and D.R.N. F.M. and P.C. received funding from the Chilean National Commission for Scientific and Technological Research (CONICYT). E.H.K. was supported by the 2017 UW-Madison Hilldale Scholarship. E.A.M. was supported by a University of Wisconsin – Madison Department of Bacteriology Fellowship. We thank the University of Wisconsin – Madison Biotechnology Center DNA Sequencing Facility for Illumina sequencing services. This research was performed in part using the Wisconsin Energy Institute computing cluster, which is supported by the Great Lakes Bioenergy Research Center as part of the U.S. Department of Energy Office of Science (DE-SC0018409).

667 **REFERENCES**

668 1. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phytogeny of
669 coexisting Prochlorococcus ecotypes. Nature 393:464–467.
670 2. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding
671 H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R,
672 Chisholm SW. 2014. Single-cell genomics reveals hundreds of coexisting
673 subpopulations in wild Prochlorococcus. Science (80-) 344:416–420.
674 3. Cohan FM. 2001. Bacterial Species and Speciation. Syst Biol 50:513–524.
675 4. N K, SE R, JW B-T, M G, R S, SW C. 2018. Fundamental differences in diversity
676 and genomic population structure between Atlantic and Pacific Prochlorococcus.
677 5. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW.
678 2006. Niche partitioning among Prochlorococcus ecotypes along ocean-scale
679 environmental gradients. Science (80-) 311:1737–1740.
680 6. Gruber-Dorninger C, Pester M, Kitzinger K, Savio DF, Loy A, Rattei T, Wagner M,
681 Daims H. 2015. Functionally relevant diversity of closely related Nitrospira in
682 activated sludge. ISME J 9:643–655.
683 7. Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species
684 challenge: Making sense of genetic and ecological diversity. Science (80-).
685 Science.
686 8. Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ
687 detection of individual microbial cells without cultivation. Microbiol Rev. American
688 Society for Microbiology.
689 9. Comte J, Fauteux L, Giorgio PA del. 2013. Links between metabolic plasticity and

690 functional redundancy in freshwater bacterioplankton communities. Front
691 Microbiol 4:112.
692 10. Hesselmann RPX, Werlen C, Hahn D, van der Meer JR, Zehnder AJB. 1999.
693 Enrichment, Phylogenetic Analysis and Detection of a Bacterium That Performs
694 Enhanced Biological Phosphate Removal in Activated Sludge. Syst Appl Microbiol
695 22:454–465.
696 11. He S, Mcmahon KD. 2011. Microbiology of “Candidatus Accumulibacter” in
697 activated sludge. Microb Biotechnol 4:603–619.
698 12. Seviour RJ, Mino T, Onuki M. 2003. The microbiology of biological phosphorus
699 removal in activated sludge systems. FEMS Microbiol Rev 27:99–127.
700 13. Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. 2016.
701 Metatranscriptomic insights on gene expression and regulatory controls in
702 Candidatus Accumulibacter phosphatis. ISME J 10:810–822.
703 14. Camejo PY, Oyserman BO, McMahon KD, Noguera DR. 2019. Integrated Omic
704 Analyses Provide Evidence that a “Candidatus Accumulibacter phosphatis” Strain
705 Performs Denitrification under Microaerobic Conditions. mSystems 4:e00193-18.
706 15. He S, Kunin V, Haynes M, Martin HG, Ivanova N, Rohwer F, Hugenholtz P,
707 McMahon KD. 2010. Metatranscriptomic array analysis of “Candidatus
708 Accumulibacter phosphatis”-enriched enhanced biological phosphorus removal
709 sludge. Environ Microbiol 12:1205–1217.
710 16. Law Y, Kirkegaard RH, Cokro AA, Liu X, Arumugam K, Xie C, Stokholm-
711 Bjerregaard M, Drautz-Moses DI, Nielsen PH, Wuertz S, Williams RBH. 2016.
712 Integrative microbial community analysis reveals full-scale enhanced biological

713 phosphorus removal under tropical conditions. Sci Rep 6:1–15.
714 17. Wang Y, Gao H, Wells G. 2020. Integrated Omics Analyses Reveal Differential
715 Gene Expression and Potential for Cooperation Between Denitrifying
716 Polyphosphate and Glycogen Accumulating Organisms. bioRxiv. bioRxiv.
717 18. Mcmahon KD, Dojka M a, Pace NR, Jenkins D, Keasling JD. 2002.
718 Polyphosphate Kinase from Activated Sludge Performing Enhanced Biological
719 Phosphorus Removal Polyphosphate Kinase from Activated Sludge Performing
720 Enhanced Biological Phosphorus Removal †. Appl Environ Microbiol 68:4971–
721 4978.
722 19. He S, Gall DL, McMahon KD. 2007. “Candidatus accumulibacter” population
723 structure in enhanced biological phosphorus removal sludges as revealed by
724 polyphosphate kinase genes. Appl Environ Microbiol 73:5865–5874.
725 20. Peterson SB, Warnecke F, Madejska J, McMahon KD, Hugenholtz P. 2008.
726 Environmental distribution and population biology of Candidatus Accumulibacter,
727 a primary agent of biological phosphorus removal. Environ Microbiol 10:2692–
728 2703.
729 21. McMahon KD, Yilmaz S, He S, Gall DL, Jenkins D, Keasling JD. 2007.
730 Polyphosphate kinase genes from full-scale activated sludge plants. Appl
731 Microbiol Biotechnol 77:167–173.
732 22. Flowers JJ, He S, Malfatti S, del Rio TG, Tringe SG, Hugenholtz P, McMahon KD.
733 2013. Comparative genomics of two “Candidatus Accumulibacter” clades
734 performing biological phosphorus removal. ISME J 7:2301–2314.
735 23. Camejo PY, Owen BR, Martirano J, Ma J, Kapoor V, Santo Domingo J, McMahon

736 KD, Noguera DR. 2016. Candidatus Accumulibacter phosphatis clades enriched
737 under cyclic anaerobic and microaerobic conditions simultaneously use different
738 electron acceptors. Water Res 102:125–137.
739 24. Martín HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C,
740 He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL,
741 Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P. 2006.
742 Metagenomic analysis of two enhanced biological phosphorus removal (EBPR)
743 sludge communities. Nat Biotechnol 24:1263–1269.
744 25. Flowers JJ, He S, Yilmaz S, Noguera DR, McMahon KD. 2009. Denitrification
745 capabilities of two biological phosphorus removal sludges dominated by different
746 "Candidatus Accumulibacter" clades. Environ Microbiol Rep 1:583–588.
747 26. Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. 2015. Expanding our
748 view of genomic diversity in *C andidatus* Accumulibacter clades. Environ
749 Microbiol 17:1574–1585.
750 27. Gao H, Mao Y, Zhao X, Liu WT, Zhang T, Wells G. 2019. Genome-centric
751 metagenomics resolves microbial diversity and prevalent truncated denitrification
752 pathways in a denitrifying PAO-enriched bioprocess. Water Res 155:275–287.
753 28. Carvalho G, Lemos PC, Oehmen A, Reis MAM. 2007. Denitrifying phosphorus
754 removal: Linking the process performance with the microbial community structure.
755 Water Res 41:4383–4396.
756 29. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming
757 tool for FastQ files.
758 30. Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to

759 improve genome assemblies. Bioinformatics 27:2957–2963.
760 31. Li H. 2013. Seqtk: a fast and lightweight tool for processing FASTA or FASTQ
761 sequences.
762 32. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly
763 using de Bruijn graphs. Genome Res 18:821–829.
764 33. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: An extension of
765 Velvet assembler to de novo metagenome assembly from short sequence reads.
766 Nucleic Acids Res 40.
767 34. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an
768 automated binning method to recover individual genomes from metagenomes
769 using an expectation-maximization algorithm. Microbiome 2:26.
770 35. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R,
771 Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–
772 51.
773 36. Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl
774 JL, Ivanova N, Woyke T, Kyrpides N, Pati A. 2016. ProDeGe: A computational
775 protocol for fully automated decontamination of genomes. ISME J 10:269–272.
776 37. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.
777 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data.
778 PeerJ 3:e1319.
779 38. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I.
780 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long
781 reads. Gigascience 4:35.

782 39. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T, Sun
783 N, Rao J, Wang J, Zhang W, Chen Y, Liao S, Jiang H, Liu X, Yang Z, Mu F, Gao
784 S. 2017. LR_Gapcloser: a tiling path-based gap closer that uses long reads to
785 complete genome assembly. Gigascience 6:1.
786 40. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
787 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler
788 G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly
789 algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–77.
790 41. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read
791 merging via overlap. PLoS One 12:e0185056.
792 42. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for
793 accurately reconstructing single genomes from complex microbial communities.
794 PeerJ 3:e1165.
795 43. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: A tool for fast and
796 accurate genomic comparisons that enables improved genome recovery from
797 metagenomes through de-replication. ISME J 11:2864–2868.
798 44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
799 assessing the quality of microbial genomes recovered from isolates, single cells,
800 and metagenomes. Genome Res 25:1043–55.
801 45. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
802 30:2068–2069.
803 46. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H.
804 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and

805 adaptive score threshold. bioRxiv 602110.
806 47. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
807 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
808 30:3059–66.
809 48. Stamatakis A. 2014. The RAxML v8 . 0 . X Manual 1–55.
810 49. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
811 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
812 boundaries. Nat Commun 9:5114.
813 50. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to
814 classify genomes with the Genome Taxonomy Database. Bioinformatics.
815 51. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the
816 display and annotation of phylogenetic and other trees. Nucleic Acids Res
817 44:W242–W245.
818 52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden
819 TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
820 53. Melnyk RA, Hossain SS, Haney CH. 2019. Convergent gain and loss of genomic
821 islands drive lifestyle changes in plant-associated Pseudomonas. ISME J
822 13:1575–1588.
823 54. Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using
824 DIAMOND. Nat Methods. Nature Publishing Group.
825 55. Sonnhammer ELL, Östlund G. 2015. InParanoid 8: Orthology analysis between
826 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43:D234–D239.
827 56. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-

828 scale detection of protein families. Nucleic Acids Res 30:1575–84.
829 57. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol
830 7:e1002195.
831 58. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and
832 Visualization of Phylogenomic Data. Mol Biol Evol 33:1635–1638.
833 59. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D,
834 Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson
835 J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ,
836 Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R,
837 Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F,
838 van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner
839 A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C,
840 Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V., Olivetti E, Martin E,
841 Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen
842 GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J,
843 Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV,
844 Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K,
845 Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ,
846 Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT,
847 Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T,
848 Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T,
849 Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. 2020. SciPy 1.0:
850 fundamental algorithms for scientific computing in Python. Nat Methods 17:261–

851 272.
852 60. Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the
853 visualization of intersecting sets and their properties. Bioinformatics 33:2938–
854 2940.
855 61. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund
856 G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K,
857 Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo
858 K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.
859 62. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
860 preprocessor. Bioinformatics 34:i884–i890.
861 63. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of
862 ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.
863 64. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010.
864 Prodigal: prokaryotic gene recognition and translation initiation site identification.
865 BMC Bioinformatics 11:119.
866 65. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-
867 seq quantification. Nat Biotechnol 34:525–527.
868 66. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
869 dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.
870 67. Oyserman BO, Moya F, Lawson CE, Garcia AL, Vogt M, Heffernen M, Noguera
871 DR, McMahon KD. 2016. Ancestral genome reconstruction identifies the
872 evolutionary basis for trait acquisition in polyphosphate accumulating bacteria.
873 ISME J 10:2931–2945.

874 68. Watkins PA. 2018. The Fatty Acyl-CoA SynthetasesReference Module in
875 Biomedical Sciences. Elsevier.
876 69. He S, McMahon KD. 2011. “Candidatus Accumulibacter” gene expression in
877 response to dynamic EBPR conditions. ISME J 5:329–340.
878 70. Long M, Liu J, Chen Z, Bleijlevens B, Roseboom W, Albracht SPJ. 2007.
879 Characterization of a HoxEFUYH type of [NiFe] hydrogenase from
880 Allochromatium vinosum and some EPR and IR properties of the hydrogenase
881 module. J Biol Inorg Chem 12:62–78.
882 71. Antal TK, Oliveira P, Lindblad P. 2006. The bidirectional hydrogenase in the
883 cyanobacterium Synechocystis sp. strain PCC 6803. Int J Hydrogen Energy
884 31:1439–1444.
885 72. Vignais PM, Billoud B. 2007. Occurrence, classification, and biological function of
886 hydrogenases: An overview. Chem Rev. Chem Rev.
887 73. Guedes da Silva L, Olavarria Gamez K, Castro Gomes J, Akkermans K, Welles L,
888 Abbas B, van Loosdrecht MCM, Wahl SA. 2020. Revealing the metabolic
889 flexibility of Candidatus Accumulibacter phosphatis through redox cofactor
890 analysis and metabolic network modeling. Appl Environ Microbiol.
891 74. Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, Hettich RL,
892 Bond PL, VerBerkmoes NC, Banfield JF. 2008. Community proteogenomics
893 highlights microbial strain-variant protein expression within activated sludge
894 performing enhanced biological phosphorus removal. ISME J 2:853–864.
895 75. Hesselmann RPX, Von Rummell R, Resnick SM, Hany R, Zehnder AJB. 2000.
896 Anaerobic metabolism of bacteria performing enhanced biological phosphate

897 removal. Water Res 34:3487–3494.
898 76. Scarborough MJ, Lawson CE, Hamilton JJ, Donohue TJ, Noguera DR. 2018.
899 Metatranscriptomic and Thermodynamic Insights into Medium-Chain Fatty Acid
900 Production Using an Anaerobic Microbiome. mSystems 3:e00221-18.
901 77. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J,
902 Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG,
903 Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR.
904 2016. Genome-wide selective sweeps and gene-specific sweeps in natural
905 bacterial populations. ISME J 10:1589–1601.
906 78. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T,
907 Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018.
908 Contrasting patterns of genome-level diversity across distinct co-occurring
909 bacterial populations. ISME J 12:742–755.
910 79. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist,
911 metagenomics reveal. Environ Microbiol 14:347–355.
912 80. Konstantinidis KT, DeLong EF. 2008. Genomic patterns of recombination, clonal
913 divergence and environment in marine microbial populations. ISME J 2:1052–
914 1065.
915 81. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield
916 JF. 2020. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial
917 Species Boundaries. mSystems 5.
918 82. Cohan FM. 2006. Towards a conceptual and operational union of bacterial
919 systematics, ecology, and evolution, p. 1985–1996. *In* Philosophical Transactions

920 of the Royal Society B: Biological Sciences. Royal Society.
921 83. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, Cordero
922 OX. 2018. Strain-level diversity drives alternative community types in millimetre-
923 scale granular biofilms. Nat Microbiol 3:1295–1303.
924 84. Albertsen M, Benedicte L, Hansen S, Saunders AM, Nielsen H, Lehmann Nielsen
925 K. 2011. A metagenome of a full-scale microbial community carrying out
926 enhanced biological phosphorus removal. ISME J 6:1094–1106.
927 85. Albertsen M, McIlroy SJ, Stokholm-Bjerregaard M, Karst SM, Nielsen PH. 2016.
928 "Candidatus Propionivibrio aalborgensis": A novel glycogen accumulating
929 organism abundant in full-scale enhanced biological phosphorus removal plants.
930 Front Microbiol 7:1033.
931 86. Welles L, Abbas B, Sorokin DY, Lopez-Vazquez CM, Hooijmans CM, van
932 Loosdrecht MCM, Brdjanovic D. 2017. Metabolic Response of "Candidatus
933 Accumulibacter Phosphatis" Clade II C to Changes in Influent P/C Ratio. Front
934 Microbiol 7:2121.
935 87. Welles L, Tian WD, Saad S, Abbas B, Lopez-Vazquez CM, Hooijmans CM, van
936 Loosdrecht MCM, Brdjanovic D. 2015. Accumulibacter clades Type I and II
937 performing kinetically different glycogen-accumulating organisms metabolisms for
938 anaerobic substrate uptake. Water Res 83:354–366.
939 88. Kim JM, Lee HJ, Lee DS, Jeon CO. 2013. Characterization of the denitrification-
940 associated phosphorus uptake properties of "Candidatusaccumulibacter
941 phosphatis" clades in sludge subjected to enhanced biological phosphorus
942 removal. Appl Environ Microbiol 79:1969–1979.

943 89. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE,
944 O'Malley MA, García Martín H, Pfleger BF, Raskin L, Venturelli OS, Weissbrodt
945 DG, Noguera DR, McMahon KD. 2019. Common principles and best practices for
946 engineering microbiomes. Nat Rev Microbiol. Nature Publishing Group.
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965

966 **FIGURE AND TABLE LEGENDS**

967

968 **Figure 1. Overview of Experimental Design and Enriched Accumulibacter Clades.**

969 **1A.** Overview of EBPR cycle and sampling scheme. Samples were collected for RNA
970 sequencing over the course of a normal EBPR cycle, with reported measurements for
971 soluble phosphorus and acetate normalized by volatile suspended solids (VSS). Samples
972 were taken at time-points 0, 31, 70, 115, 150, 190, and 250 minutes. **1B.** Abundance of
973 different Accumulibacter clades by *ppk1* copies per ng of DNA. Quantification of
974 Accumulibacter clades was performed by qPCR of the *ppk1* gene as described in Camejo
975 et al. (23).

976

977 **Figure 2. Assembly of Four High-Quality Accumulibacter Genomes**

978 **2A.** Phylogenetic tree of *ppk1* nucleotide sequences from available Accumulibacter
979 reference genomes, clone sequences, and the four assembled Accumulibacter genomes
980 from this study. Tips in bold represent *ppk1* sequences from metagenome-assembled
981 genomes, whereas others are sequences from clone fragments available at the
982 referenced accession numbers. Starred tips represent *ppk1* sequences from the four
983 assembled Accumulibacter genomes from this study. Tree was constructed using RaXML
984 with 100 rapid bootstraps. **2B.** Pairwise genome-wide ANI comparisons of all publicly
985 available Accumulibacter references and the four new Accumulibacter genomes, denoted
986 in bold. Genomes are grouped according to clade designation by *ppk1* sequence identity
987 and in order of the *ppk1* phylogeny, except for those that do not fall in an established
988 clade. **2C.** Genome diagram of the UW6 Accumulibacter IIC MAG (98.6% completeness,

989 2.5% contamination, 5.18 Mbp). Layers represent from top to bottom: coverage over 1000
990 bp windows, GC content and skew across the same 1000 bp windows, coding sequences
991 on the positive and negative strands, and positions of predicted rRNA (red) and tRNA
992 (black) sequences.
993
994 **Figure 3. Core and Flexible Gene Content among the Accumulibacter lineage and**
995 **UW genomes**
996 **3A.** Phylogenetic tree of the coding regions for the *ppk1* locus among high-quality
997 Accumulibacter references, UW genomes, and outgroup taxa *Dechloromonas aromatica*
998 *RCB* and *Rhodocyclus tenuis DSM 110.* The pie graph at each node represents the
999 fraction of genes that are shared and flexible for genomes belonging to that clade. **3B.**
1000 Upset plot representing the intersections of groups of orthologous gene clusters among
1001 UW genomes and outgroups, analogous to a Venn diagram. 1075 gene families were
1002 shared among all UW Accumulibacter clades and outgroup genomes and 277 separate
1003 gene families were found only within the UW Accumulibacter clades. Each bar represents
1004 the number of orthologous gene clusters, and the dot plot represents the genomes in
1005 which the groups intersect.
1006
1007 **Figure 4. Transcriptomic Profiles of Accumulibacter Clades**
1008 **4A.** Average expression profiles of each reference genome in the anaerobic and aerobic
1009 phases. The IIC-UW6 genome mapped the most transcripts even though the IA-UW4 was
1010 more abundant. Reads were competitively mapped to the four assembled Accumulibacter
1011 clade genomes and normalized by transcripts per million (TPM). The sum of the

1012 normalized counts of the three samples in the anaerobic phase and four samples in the
1013 aerobic phase were averaged, respectively, and plotted on a log_{10} scale. Numbers and
1014 percentages for adjacent to normalized bars represent the raw number of reads that are
1015 mapped to each genome. **4B.** Top 50 differentially-expressed genes in clade IIC-UW6.
1016 All 50 genes pass the threshold of ±1.5 log-10 fold-change across the cycle. Annotations
1017 were predicted from a combination of Prokka (45) and KofamKOALA (46). **4C.**
1018 Denitrification gene expression among clades IIC-UW6, IA-UW4, and IIA-UW5. We
1019 summarized expression for genes involved in denitrification by highlighting parts of the
1020 cycle in which certain genes are expressed the most. This is characterized by acetate
1021 contact (pink), early anaerobic (dark purple), late anaerobic (purple), early aerobic (dark
1022 blue) and late aerobic (light blue), where uniform describes uniform expression across
1023 the entire cycle (grey). If an arrow is missing for a step, that particular clade did not contain
1024 confident annotations within the genome for those genes. For nitrate reduction, the *nar*
1025 system is represented as a black dot and the *nap* system is represented as a grey dot.
1026 Multiple arrows are shown for a single step in which particular genes were highly
1027 expressed in multiple phases.
1028
1029 **Figure 5. Differential Expression of Shared Genes in Strains IA-UW4 and IIC-UW6**
1030 Gene expression profiles of core COGs across the EBPR cycle that are differentially
1031 expressed between anaerobic and aerobic conditions in either IA-UW4 or IIC-UW6 or
1032 both strains. To consider a gene differentially expressed in a strain, the COG must exhibit
1033 ±1.5 fold-change in expression between anaerobic and aerobic phases. Homologs that

1034 are directly compared to each other in IA-UW4 and IIC-UW6 are orthologous genes that
1035 belong to the same cluster.
1036
1037 **Figure 6. Differential Expression of Accessory Genes in Strain IA-UW4**
1038 Gene expression profiles of groups of accessory gene groups in strain IA-UW4 that are
1039 not present in the IIC-UW6 genome. A gene was considered differentially expressed
1040 between the anaerobic and aerobic conditions if it exhibited greater or less than ±1.5 fold-
1041 change in expression. Colored dots represent whether or not that gene contains orthologs
1042 within outgroup genomes, other Type II Accumulibacter other than the IIC-UW6 genome,
1043 within Type I Accumulibacter, or only within IA.
1044
1045 **Table 1. Assembled Accumulibacter Genome Statistics**
1046 Genome quality calculations for completeness and contaminations were made with
1047 CheckM based on the presence/absence of single-copy genes (44).
1048
1049 **Table 2. Metatranscriptomic Reads Mapped to Accumullibacter Genomes**
1050 Counts of raw metatranscriptomic reads mapped to each Accumulibacter genome with
1051 kallisto (65) in each sample for the three anaerobic samples and four aerobic samples
1052 and total reads mapped for all samples.
1053
1054 **Supplementary Figure 1.**
1055 Pairwise Pearson correlation coefficient of shared gene content as a measure of how
1056 similar each of the UW genomes is to each other based on COGs.

1057 **Supplementary Figure 2.**

1058 **S2A)** Species tree of all Accumulibacter reference genomes, UW-assembled genomes,
1059 and outgroup genomes using single-copy markers from the GTDB-tk (50). A phylogenetic
1060 tree was constructed using RAxmL with 100 rapid bootstraps and visualized in iTOL. **S2B)**
1061 Pairwise BLAST percent identity scores for *ppk1* coding regions for each reference
1062 genome in Figure 2B, in order of the *ppk1* phylogeny in Figure 2A. Raw pairwise percent
1063 identity scores are available at
1064 https://figshare.com/articles/dataset/Pairwise_Accumulibacter_ppk1_BLAST_Percent_D
1065 _Scores/13237247.

1066

1067 **Supplementary Table 1**

1068 Genome accessions and statistics for all Accumulibacter references used in phylogenies,
1069 ANI comparisons, and COG calculations. A reference genome was used for COG
1070 comparisons if it was a high quality genome (>95% complete and <5% redundant).
1071 Available at https://figshare.com/account/projects/90614/articles/13237148

1072

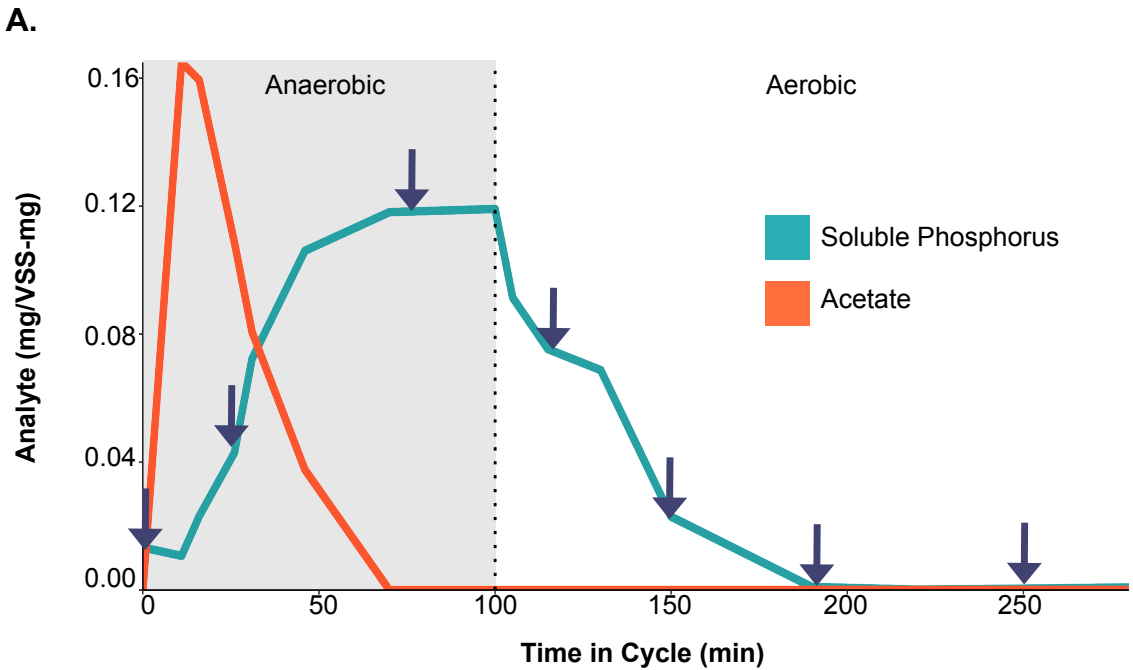
1073 **Supplementary File 1**

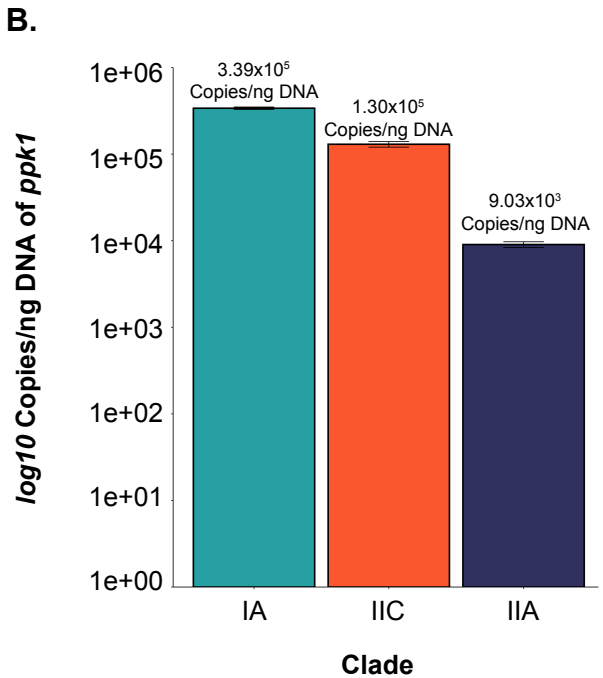
1074 Raw pairwise genome-wide ANI results for each genome calculated with FastANI.
1075 Available at https://figshare.com/account/projects/90614/articles/13237244

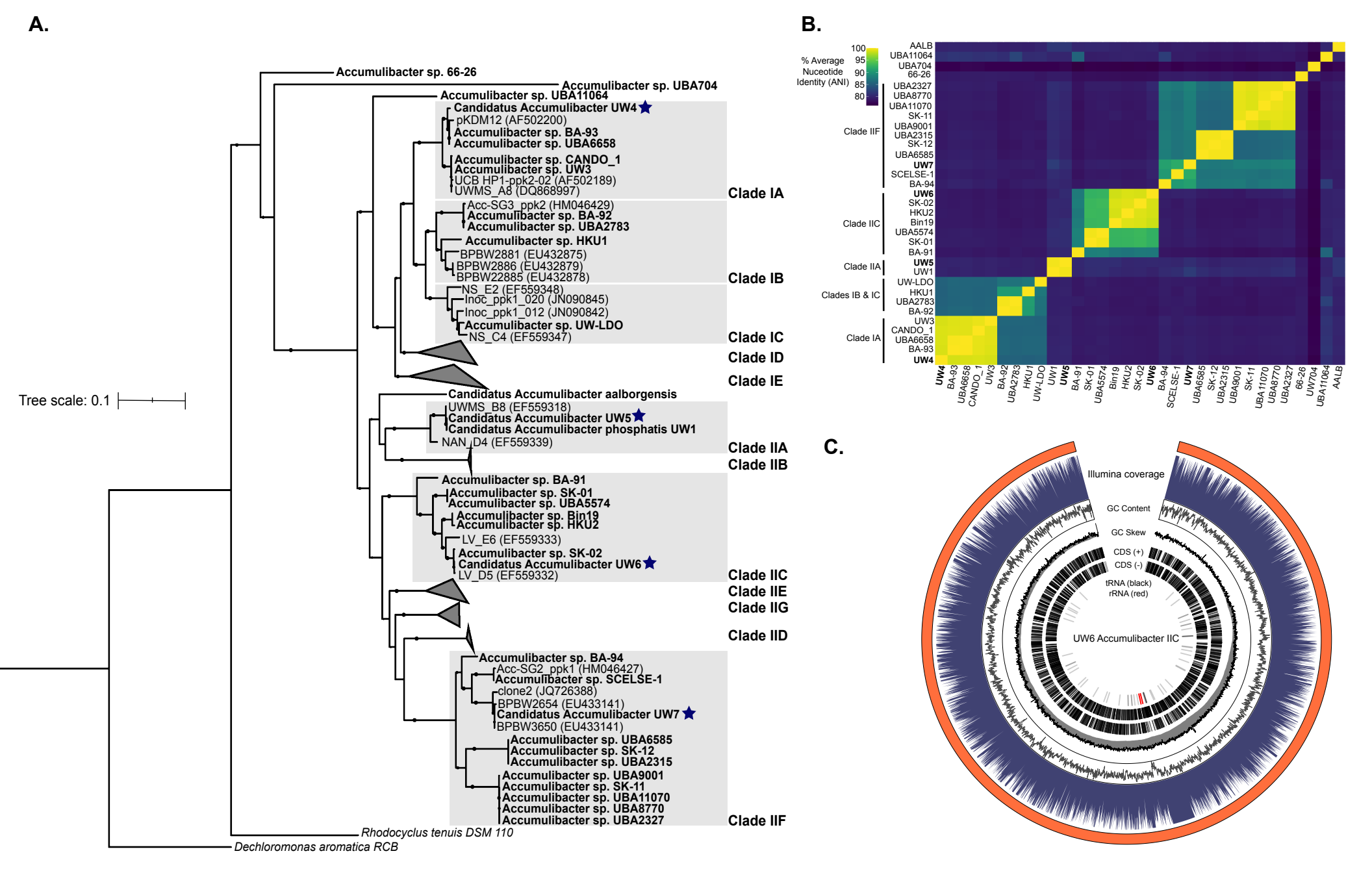
1076

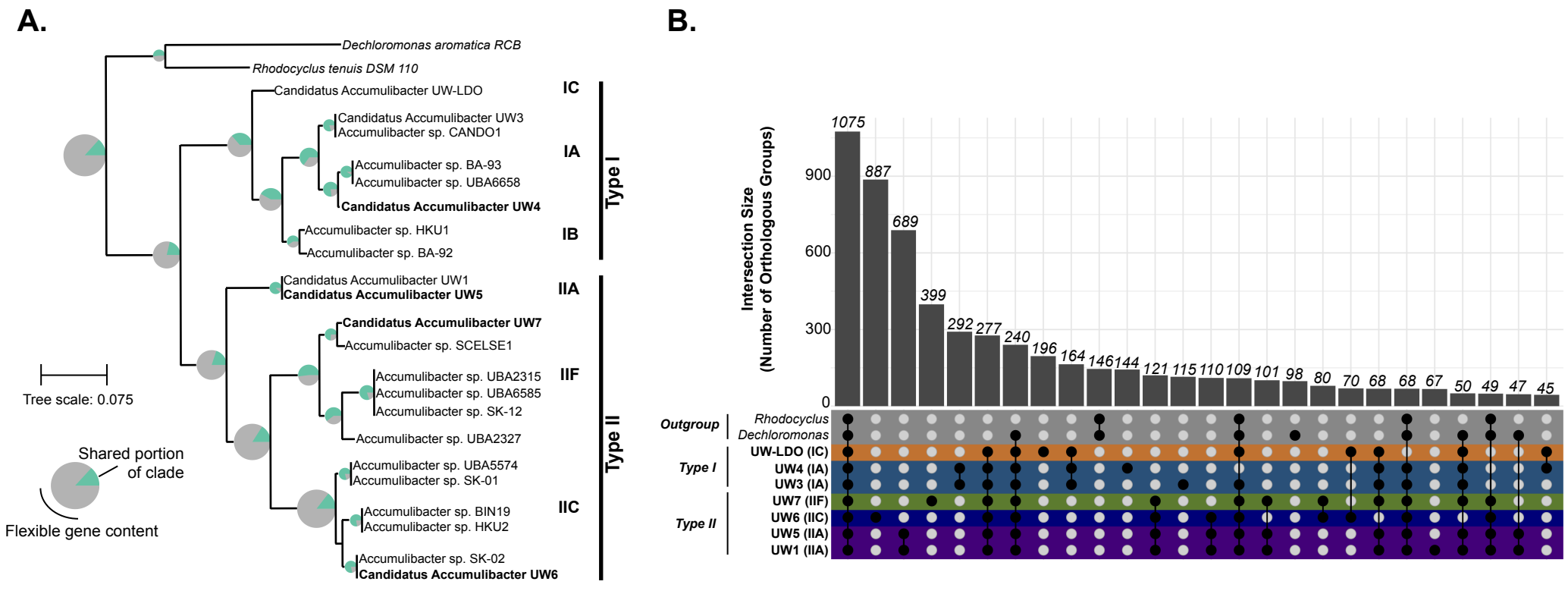
1077 **Supplementary File 2**

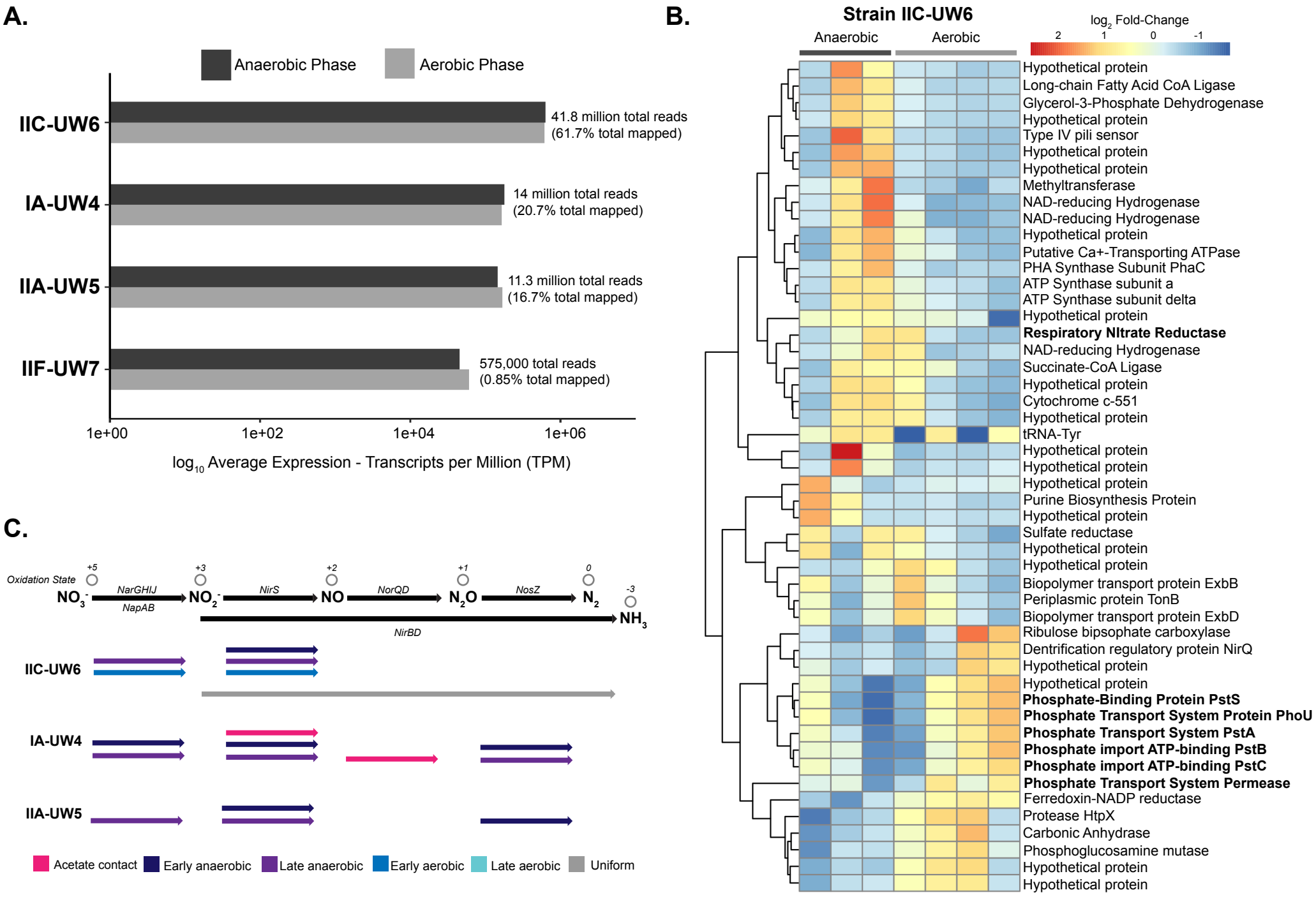
1078 Raw pairwise BLAST percent identity (PID) results for *ppk1* genes from each reference
1079 genome. Available at https://figshare.com/account/projects/90614/articles/13237247

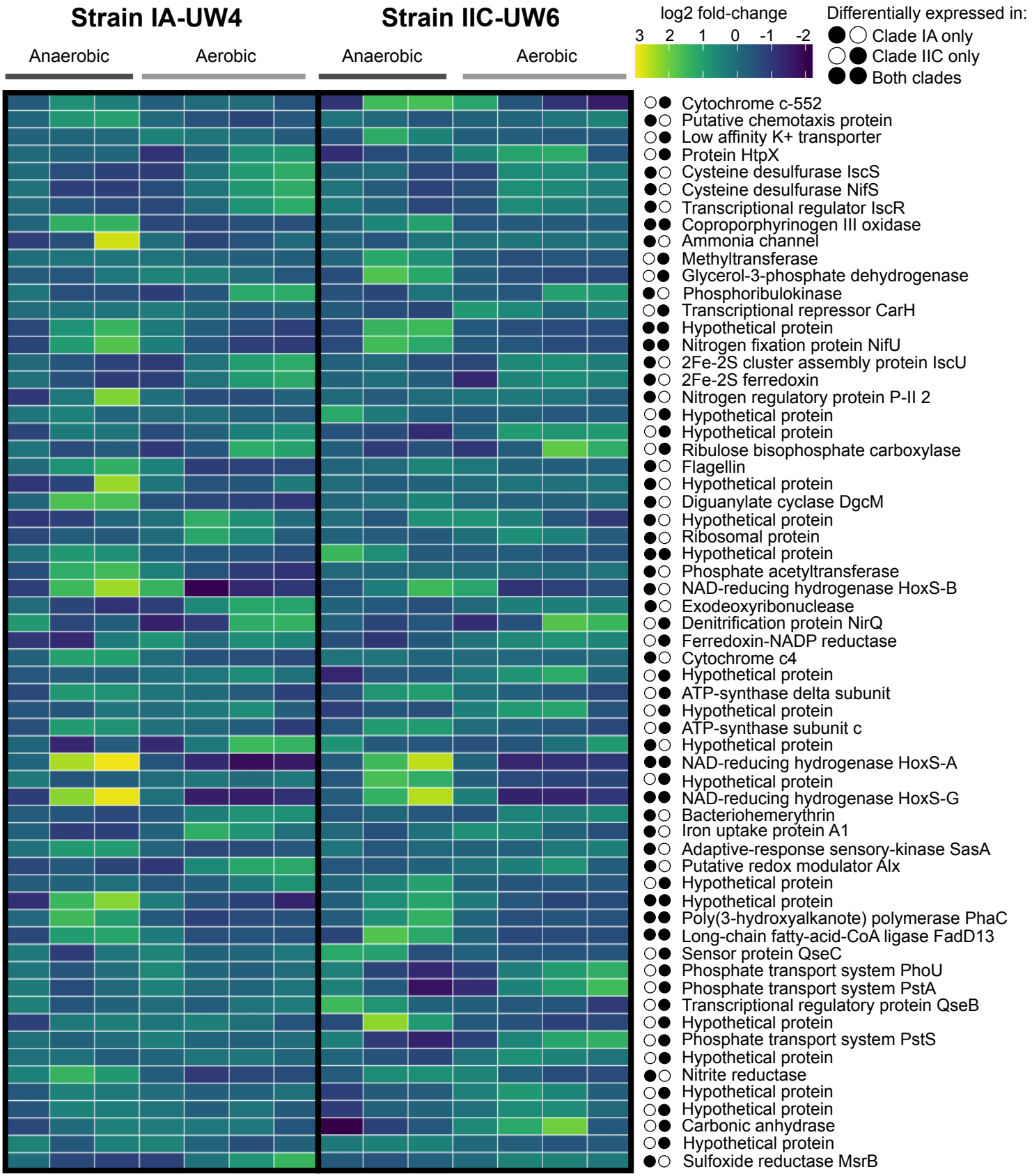

A.
Anaerobic
Aerobic
Soluble Phosphorus
Acetate
Analyte (mg/VSS-mg)
0.16
0.12
0.08
0.04
0.00
0
50
100
150
200
250
Time in Cycle (min)

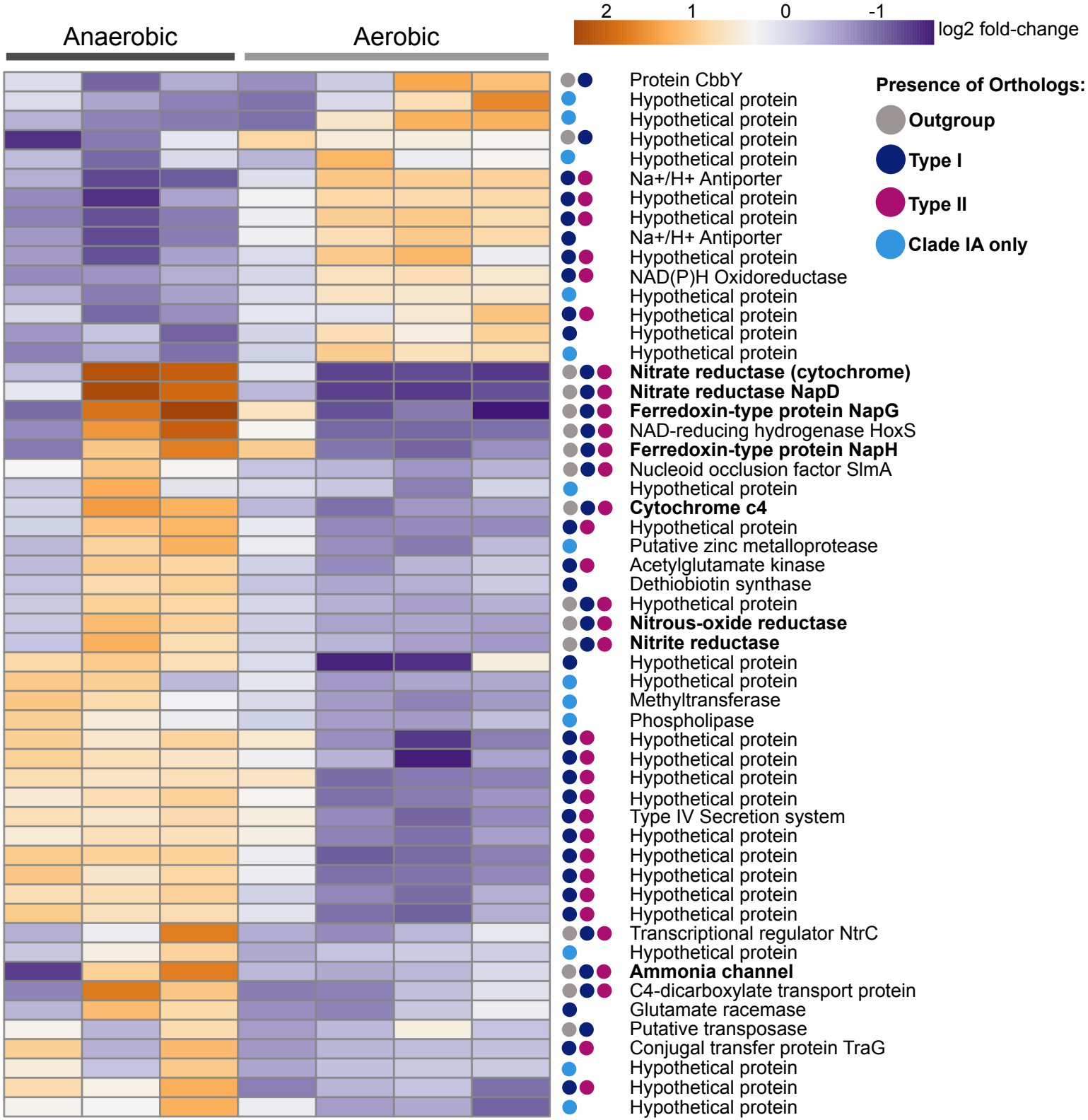

B.
3.39x10^5
Copies/ng DNA
1.30x10^5
Copies/ng DNA
9.03x10^3
Copies/ng DNA
log10 Copies/ng DNA of ppk1
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1e+00
IA
IIC
IIA
Clade


A.
Accumulibacter sp. 66-26
Accumulibacter sp. UBA704
Accumulibacter sp. UBA11064
Candidatus Accumulibacter UW4
pKDM12 (AF502200)
Accumulibacter sp. BA-93
Accumulibacter sp. UBA6658
Accumulibacter sp. CANDO_1
Accumulibacter sp. UW3
UCB HP1-ppk2-02 (AF502189)
UWMS_A8 (DQ868997)
Clade IA
Acc-SG3_ppk2 (HM046429)
Accumulibacter sp. BA-92
Accumulibacter sp. UBA2783
Accumulibacter sp. HKU1
BPBW2881 (EU432875)
BPBW2886 (EU432879)
BPBW22885 (EU432878)
Clade IB
NS_E2 (EF559348)
Inoc_ppk1_020 (JN090845)
Inoc_ppk1_012 (JN090842)
Accumulibacter sp. UW-LDO
NS_C4 (EF559347)
Clade IC
Clade ID
Clade IE
Tree scale: 0.1
Candidatus Accumulibacter aalborgensis
UWMS_B8 (EF559318)
Candidatus Accumulibacter UW5
Candidatus Accumulibacter phosphatis UW1
NAN_D4 (EF559339)
Clade IIA
Clade IIB
Accumulibacter sp. BA-91
Accumulibacter sp. SK-01
Accumulibacter sp. UBA5574
Accumulibacter sp. Bin19
Accumulibacter sp. HKU2
LV_E6 (EF559333)
Accumulibacter sp. SK-02
Candidatus Accumulibacter UW6
LV_D5 (EF559332)
Clade IIC
Clade IIE
Clade IIG
Clade IID
Accumulibacter sp. BA-94
Acc-SG2_ppk1 (HM046427)
Accumulibacter sp. SCELSE-1
clone2 (JQ726388)
BPBW2654 (EU433141)
Candidatus Accumulibacter UW7
BPBW3650 (EU433141)
Accumulibacter sp. UBA6585
Accumulibacter sp. SK-12
Accumulibacter sp. UBA2315
Accumulibacter sp. UBA9001
Accumulibacter sp. SK-11
Accumulibacter sp. UBA11070
Accumulibacter sp. UBA8770
Accumulibacter sp. UBA2327
Clade IIF
Rhodocyclus tenuis DSM 110
Dechloromonas aromatica RCB
B.
% Average Nuceotide Identity (ANI)
100
95
90
85
80
Clade IIF
Clade IIC
Clade IIA
Clades IB & IC
Clade IA
AALB
UBA11064
UBA704
66-26
UBA2327
UBA8770
UBA11070
SK-11
UBA9001
UBA2315
SK-12
UBA6585
UW7
SCELSE-1
BA-94
UW6
SK-02
HKU2
Bin19
UBA5574
SK-01
BA-91
UW5
UW1
UW-LDO
HKU1
UBA2783
BA-92
UW3
CANDO_1
UBA6658
BA-93
UW4
C.
Illumina coverage
GC Content
GC Skew
CDS (+)
CDS (-)
tRNA (black)
rRNA (red)
UW6 Accumulibacter IIC


A.
Dechloromonas aromatica RCB
Rhodocyclus tenuis DSM 110
Candidatus Accumulibacter UW-LDO
IC
Candidatus Accumulibacter UW3
Accumulibacter sp. CANDO1
IA
Accumulibacter sp. BA-93
Accumulibacter sp. UBA6658
Candidatus Accumulibacter UW4
Accumulibacter sp. HKU1
IB
Accumulibacter sp. BA-92
Type I
Candidatus Accumulibacter UW1
Candidatus Accumulibacter UW5
IIA
Candidatus Accumulibacter UW7
Accumulibacter sp. SCELSE1
Accumulibacter sp. UBA2315
Accumulibacter sp. UBA6585
Accumulibacter sp. SK-12
IIF
Accumulibacter sp. UBA2327
Accumulibacter sp. UBA5574
Accumulibacter sp. SK-01
Accumulibacter sp. BIN19
Accumulibacter sp. HKU2
IIC
Accumulibacter sp. SK-02
Candidatus Accumulibacter UW6
Type II
Tree scale: 0.075
Shared portion of clade
Flexible gene content
B.
Intersection Size (Number of Orthologous Groups)
0
300
600
900
1075
887
689
399
292
277
240
196
164
146
144
121
115
110
109
101
98
80
70
68
68
67
50
49
47
45
Outgroup
Rhodocyclus
Dechloromonas
Type I
UW-LDO (IC)
UW4 (IA)
UW3 (IA)
Type II
UW7 (IIF)
UW6 (IIC)
UW5 (IIA)
UW1 (IIA)

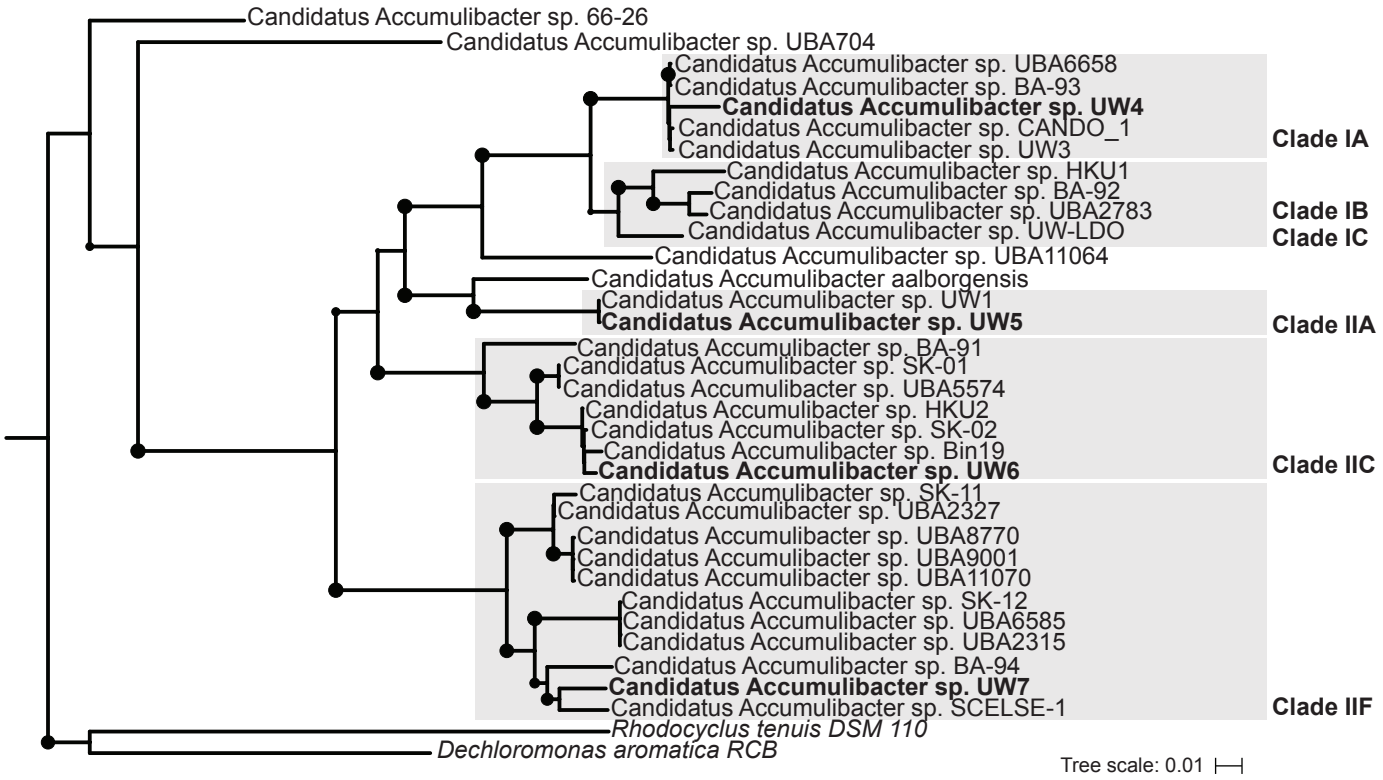

A.
Anaerobic Phase
Aerobic Phase
IIC-UW6
41.8 million total reads
(61.7% total mapped)
IA-UW4
14 million total reads
(20.7% total mapped)
IIA-UW5
11.3 million total reads
(16.7% total mapped)
IIF-UW7
575,000 total reads
(0.85% total mapped)
1e+00
1e+02
1e+04
1e+06
log10 Average Expression - Transcripts per Million (TPM)
B.
Strain IIC-UW6
Anaerobic
Aerobic
log2 Fold-Change
2
1
0
-1
Hypothetical protein
Long-chain Fatty Acid CoA Ligase
Glycerol-3-Phosphate Dehydrogenase
Hypothetical protein
Type IV pili sensor
Hypothetical protein
Hypothetical protein
Methyltransferase
NAD-reducing Hydrogenase
NAD-reducing Hydrogenase
Hypothetical protein
Putative Ca+-Transporting ATPase
PHA Synthase Subunit PhaC
ATP Synthase subunit a
ATP Synthase subunit delta
Hypothetical protein
Respiratory NItrate Reductase
NAD-reducing Hydrogenase
Succinate-CoA Ligase
Hypothetical protein
Cytochrome c-551
Hypothetical protein
tRNA-Tyr
Hypothetical protein
Hypothetical protein
Hypothetical protein
Purine Biosynthesis Protein
Hypothetical protein
Sulfate reductase
Hypothetical protein
Hypothetical protein
Biopolymer transport protein ExbB
Periplasmic protein TonB
Biopolymer transport protein ExbD
Ribulose bipsophate carboxylase
Dentrification regulatory protein NirQ
Hypothetical protein
Hypothetical protein
Phosphate-Binding Protein PstS
Phosphate Transport System Protein PhoU
Phosphate Transport System PstA
Phosphate import ATP-binding PstB
Phosphate import ATP-binding PstC
Phosphate Transport System Permease
Ferredoxin-NADP reductase
Protease HtpX
Carbonic Anhydrase
Phosphoglucosamine mutase
Hypothetical protein
Hypothetical protein
C.
Oxidation State
+5
+3
+2
+1
0
-3
NO3-
NarGHIJ
NapAB
NO2-
NirS
NO
NorQD
N2O
NosZ
N2
NirBD
NH3
IIC-UW6
IA-UW4
IIA-UW5
Acetate contact
Early anaerobic
Late anaerobic
Early aerobic
Late aerobic
Uniform


Strain IA-UW4
Strain IIC-UW6
Anaerobic
Aerobic
Anaerobic
Aerobic
log2 fold-change
3 2 1 0 -1 -2
Differentially expressed in:
Clade IA only
Clade IIC only
Both clades
Cytochrome c-552
Putative chemotaxis protein
Low affinity K+ transporter
Protein HtpX
Cysteine desulfurase IscS
Cysteine desulfurase NifS
Transcriptional regulator IscR
Coproporphyrinogen III oxidase
Ammonia channel
Methyltransferase
Glycerol-3-phosphate dehydrogenase
Phosphoribulokinase
Transcriptional repressor CarH
Hypothetical protein
Nitrogen fixation protein NifU
2Fe-2S cluster assembly protein IscU
2Fe-2S ferredoxin
Nitrogen regulatory protein P-II 2
Hypothetical protein
Hypothetical protein
Ribulose bisophosphate carboxylase
Flagellin
Hypothetical protein
Diguanylate cyclase DgcM
Hypothetical protein
Ribosomal protein
Hypothetical protein
Phosphate acetyltransferase
NAD-reducing hydrogenase HoxS-B
Exodeoxyribonuclease
Denitrification protein NirQ
Ferredoxin-NADP reductase
Cytochrome c4
Hypothetical protein
ATP-synthase delta subunit
Hypothetical protein
ATP-synthase subunit c
Hypothetical protein
NAD-reducing hydrogenase HoxS-A
Hypothetical protein
NAD-reducing hydrogenase HoxS-G
Bacteriohemerythrin
Iron uptake protein A1
Adaptive-response sensory-kinase SasA
Putative redox modulator Alx
Hypothetical protein
Hypothetical protein
Poly(3-hydroxyalkanote) polymerase PhaC
Long-chain fatty-acid-CoA ligase FadD13
Sensor protein QseC
Phosphate transport system PhoU
Phosphate transport system PstA
Transcriptional regulatory protein QseB
Hypothetical protein
Phosphate transport system PstS
Hypothetical protein
Nitrite reductase
Hypothetical protein
Hypothetical protein
Carbonic anhydrase
Hypothetical protein
Sulfoxide reductase MsrB

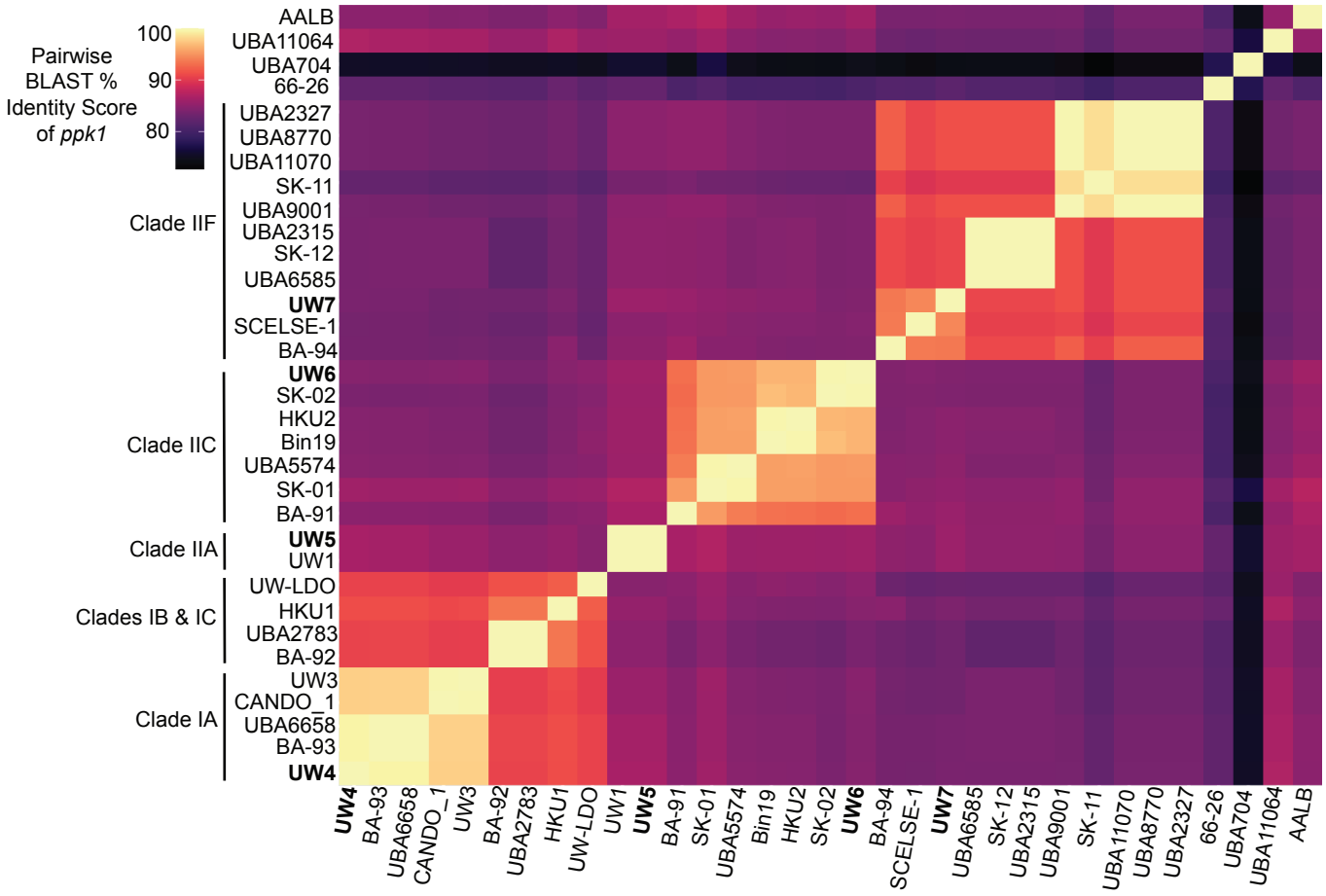

Anaerobic
Aerobic
2
1
0
-1
log2 fold-change
Presence of Orthologs:
Outgroup
Type I
Type II
Clade IA only
Protein CbbY
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein
NAD(P)H Oxidoreductase
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Nitrate reductase (cytochrome)
Nitrate reductase NapD
Ferredoxin-type protein NapG
NAD-reducing hydrogenase HoxS
Ferredoxin-type protein NapH
Nucleoid occlusion factor SlmA
Hypothetical protein
Cytochrome c4
Hypothetical protein
Putative zinc metalloprotease
Acetylglutamate kinase
Dethiobiotin synthase
Hypothetical protein
Nitrous-oxide reductase
Nitrite reductase
Hypothetical protein
Hypothetical protein
Methyltransferase
Phospholipase
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Type IV Secretion system
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Transcriptional regulator NtrC
Hypothetical protein
Ammonia channel
C4-dicarboxylate transport protein
Glutamate racemase
Putative transposase
Conjugal transfer protein TraG
Hypothetical protein
Hypothetical protein
Hypothetical protein

	UW3-IA	UW4-IA	UW-LDO-IC	UW5-IIA	UW1-IIA	UW6-IIC	UW7-IIF
UW3-IA	100.0%	78.3%	62.9%	40.9%	40.4%	33.3%	48.8%
UW4-IA	78.3%	100.0%	57.2%	34.0%	33.3%	29.2%	41.5%
UW-LDO-IC	62.9%	57.2%	100.0%	40.0%	40.0%	35.7%	48.7%
UW5-IIA	40.9%	34.0%	40.0%	100.0%	92.3%	35.7%	52.0%
UW1-IIA	40.4%	33.3%	40.0%	92.3%	100.0%	36.0%	50.1%
UW6-IIC	33.3%	29.2%	35.7%	35.7%	36.0%	100.0%	40.9%
UW7-IIF	48.8%	41.5%	48.7%	52.0%	50.1%	40.9%	100.0%

A.


Candidatus Accumulibacter sp. 66-26
Candidatus Accumulibacter sp. UBA704
Candidatus Accumulibacter sp. UBA6658
Candidatus Accumulibacter sp. BA-93
Candidatus Accumulibacter sp. UW4
Candidatus Accumulibacter sp. CANDO_1
Candidatus Accumulibacter sp. UW3
Clade IA
Candidatus Accumulibacter sp. HKU1
Candidatus Accumulibacter sp. BA-92
Candidatus Accumulibacter sp. UBA2783
Candidatus Accumulibacter sp. UW-LDO
Clade IB
Clade IC
Candidatus Accumulibacter sp. UBA11064
Candidatus Accumulibacter aalborgensis
Candidatus Accumulibacter sp. UW1
Candidatus Accumulibacter sp. UW5
Clade IIA
Candidatus Accumulibacter sp. BA-91
Candidatus Accumulibacter sp. SK-01
Candidatus Accumulibacter sp. UBA5574
Candidatus Accumulibacter sp. HKU2
Candidatus Accumulibacter sp. SK-02
Candidatus Accumulibacter sp. Bin19
Candidatus Accumulibacter sp. UW6
Clade IIC
Candidatus Accumulibacter sp. SK-11
Candidatus Accumulibacter sp. UBA2327
Candidatus Accumulibacter sp. UBA8770
Candidatus Accumulibacter sp. UBA9001
Candidatus Accumulibacter sp. UBA11070
Candidatus Accumulibacter sp. SK-12
Candidatus Accumulibacter sp. UBA6585
Candidatus Accumulibacter sp. UBA2315
Candidatus Accumulibacter sp. BA-94
Candidatus Accumulibacter sp. UW7
Candidatus Accumulibacter sp. SCELSE-1
Clade IIF
Rhodocyclus tenuis DSM 110
Dechloromonas aromatica RCB
Tree scale: 0.01

B.


Pairwise BLAST % Identity Score of ppk1
100
90
80
AALB
UBA11064
UBA704
66-26
UBA2327
UBA8770
UBA11070
SK-11
UBA9001
UBA2315
SK-12
UBA6585
UW7
SCELSE-1
BA-94
Clade IIF
UW6
SK-02
HKU2
Bin19
UBA5574
SK-01
BA-91
Clade IIC
UW5
UW1
Clade IIA
UW-LDO
HKU1
UBA2783
BA-92
Clades IB & IC
UW3
CANDO_1
UBA6658
BA-93
UW4
Clade IA
UW4
BA-93
UBA6658
CANDO_1
UW3
BA-92
UBA2783
HKU1
UW-LDO
UW1
UW5
BA-91
SK-01
UBA5574
Bin19
HKU2
SK-02
UW6
BA-94
SCELSE-1
UW7
UBA6585
SK-12
UBA2315
UBA9001
SK-11
UBA11070
UBA8770
UBA2327
66-26
UBA704
UBA11064
AALB

Genome	Completion	Contamination	Size_Mbp	Contigs	GC
IIA-UW5	98.99	5.24	4.88	68	64.3
IIC-UW6	98.57	2.46	5.18	1	61.1
IIF-UW7	98.97	3.04	4.87	102	66.3
IA-UW4	92.90	3.37	4.29	358	64.0

Clade Genome	Anaerobic-10:45	Anaerobic-11:16	Anaerobic-11:55	Aerobic-12:40	Aerobic-13:15	Aerobic-13:55	Aerobic-14:55	Total Reads Mapped
IIA-UW5	1063676	1357480	1706229	2197556	1763359	1171470	2065179	11324949
IIC-UW6	5226724	5892548	6755519	7902549	5004518	6914956	4132969	41829784
IIF-UW7	51038	78249	84520	121226	86070	90567	63866	575537
IA-UW4	1484341	2216299	2290441	3362431	1950259	1558605	1157345	14019720