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 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.

46 **IMPORTANCE**

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

60

61

62

63

64

65

66

67

68

69 **INTRODUCTION**

70 Naturally occurring microbial assemblages are composed of closely related strains
71 or subpopulations and harbor extensive genetic diversity. Coherent lineages are often
72 comprised of distinct clades, ecotypes, or “backbone subpopulations” defined by high
73 within-group sequence similarities that translate to ecologically relevant diversity (1–3).
74 For example, genetically diverse ecotypes of the marine cyanobacterium
75 *Prochlorococcus* that are 97% identical by their 16S ribosomal RNA gene sequences
76 exhibit markedly different light-dependent physiologies and distinct seasonal and
77 geographical patterns (1, 2, 4, 5). Transcriptional-level differences between closely
78 related microbial lineages have also been demonstrated, such as that of high-light and
79 low-light *Prochlorococcus* ecotypes (6). However, dissecting the emergent ecological
80 properties of genetic diversity observed among coherent microbial lineages is challenging
81 given that these principles are likely not universal among the breadth of microbial
82 diversity, and is further hindered when there is a lack of pure cultures (7–9).

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

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

105 Accumulibacter can be subdivided into two main types (I and II) and further
106 subdivided into multiple clades based upon polyphosphate kinase (*ppk1*) sequence
107 identity since the 16S rRNA marker is too highly conserved among the breadth of known
108 Accumulibacter lineages to resolve species-level differentiation (18–21). The two main
109 types share approximately 85% nucleotide identity across the *ppk1* locus (21) which
110 mirrors genome-wide average nucleotide identity (ANI) (22). Members of different clades
111 have been hypothesized to vary in their acetate uptake rates, nitrate reduction capacity,
112 and phage defense mechanisms (22, 23). For example, clade IA seems to exhibit higher
113 acetate uptake rates and phosphate release rates, whereas evidence exists that clade
114 IIA can reduce nitrate but clade IA cannot (24, 25). Additionally, the Accumulibacter

115 lineage exhibits specific trait variation within individual *ppk1*-defined clades for specific
116 traits, such as denitrification capabilities. For example, differences among clade IC strains
117 with respect to anoxic nitrate reduction coupled to phosphorus uptake suggests that the
118 ability to carry out each sequential step in denitrification could be strain-specific (14, 17,
119 26–28). The persistence of several phylogenetically distinct clades within the
120 Accumulibacter lineage in both full-scale wastewater treatment plants and lab-scale
121 enrichment bioreactors suggests that different clades may inhabit distinct ecological
122 niches (17, 29). This is largely supported by comparative genomics of Accumulibacter
123 metagenome-assembled genomes (MAGs) recovered from different bioreactor systems
124 and classified using the *ppk1* locus (14, 17, 22–24, 28, 30). However, the ecological roles
125 that distinct Accumulibacter clades may play are not yet clear.

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

138 **MATERIALS AND METHODS**

139 **Bioreactor Operation and Sample Collection**

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

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

161 chloroform separation and RNA precipitation. RNA was purified following an on-column
162 DNAse digestion using the RNase-Free DNase Set (Qiagen, Venlo, Netherlands) and
163 cleaned up with the RNeasy Mini Kit (Qiagen, Venlo, Netherlands). Accumulibacter clade
164 quantification was performed on biomass samples collected on the same day of the
165 metatranscriptomics experiment using clade-specific *ppk1* qPCR primers as described by
166 Camejo et al. (29).

167 **Metagenomic and Metatranscriptomic Library Construction and Sequencing**

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

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

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

198 **Metagenomic Assembly and Annotation**

199 We applied two different assembly and binning approaches to obtain high-quality
200 genomes of the dominant Accumulibacter strains present in the reactors. For the first
201 approach, Illumina unmerged reads were quality filtered and trimmed using the Sickle
202 software v1.33 (31). Reads were merged with FLASH v1.0.3 22 (32), with a mismatch
203 value of ≤0.25 and a minimum of ten overlapping bases from paired sequences, resulting
204 in merged read lengths of 150 to 290 bp. FASTQ files were then converted to FASTA
205 format using the Seqtk software v1.0 (33). Metagenomic reads from all three samples
206 were then co-assembled using the Velvet assembler with a k-mer size of 65 bp, a

207 minimum contig length of 200 bp, and a paired-end insert size of 300 bp (34). Metavelvet
208 was used to improve the assembly generated by Velvet (35). Metagenomic contigs were
209 then binned using Maxbin (36). This approach allowed us to assemble high-quality
210 genomes of clades IIA and IIC, named UW5 and UW6, respectively. Long Nanopore
211 reads were generated under the standard protocol and used to scaffold assembled
212 contigs with MeDuSa (37), and manually inspected to remove short contigs and
213 decontaminated using ProDeGe (38) and Anvi'o (39). Further scaffolding was performed
214 on both of these genomes using Nanopore long reads using LINKS (40) and GapCloser
215 (41).

216 Because this pipeline did not produce high-quality genomes of clades IA and IIF,
217 we applied a different approach to assemble these genomes. Quality filtered reads from
218 each metagenome were individually and co-assembled assembled using metaSPAdes
219 (42). Reads from each metagenome were mapped against each of the assemblies using
220 bbmap with a 95% sequence identity cutoff to obtain differential coverage (43), and
221 contigs were binned using MetaBat (44). Identical clusters of bins were dereplicated
222 across assemblies using dRep (45) to obtain the best quality genome of both clades IA
223 and IIF, named UW4 and UW7, respectively. All genomes were quality checked using
224 CheckM (46) and functional annotations assigned with Prokka and KofamKOALA (47,
225 48).

226 Each genome was assigned to a particular clade both by comparing the *ppk1*
227 sequence identity and ANI to previously published Accumulibacter genomes. A *ppk1*
228 database was created using sequences from previously published Accumulibacter
229 references and clone sequences. We searched for the *ppk1* gene in our draft genomes

230 using this database and aligned the corresponding hits with MAFFT (49). A phylogenetic
231 tree of aligned *ppk1* gene sequences from Accumulibacter references, select clone
232 sequences, and the outgroups *Dechloromonas aromatica RCB* and *Rhodoc*yclus *tenuis*
233 *DSM 110* with RAxML v8.1 with 100 rapid bootstraps (50). To use any given
234 Accumulibacter reference genome in downstream comparisons, a confident *ppk1* hit had
235 to be identified using the above methods. As a technical note, the IA-UW2 strain
236 assembled by Flowers et al. (22) was renamed to UW3, as the original assembly
237 contained an additional contig that was likely from a prophage. The prophage contig was
238 removed from the IA-UW2 strain and renamed to UW3, and therefore the UW numerical
239 nomenclature follows in sequential order after UW3. A phylogenetic tree of *ppk1*
240 nucleotide sequences was performed and compared to select clone sequences. Pairwise
241 genome-wide ANI was calculated between the four Accumulibacter draft genomes and
242 all available Accumulibacter genomes in Genbank with FastANI (51). A species tree of all
243 Accumulibacter references and outgroup genomes was constructed using single-copy
244 markers from the GTDBK-tk (52), aligned with MAFFT (49), built with RAxmL with 100
245 rapid bootstraps (50), and visualized in iTOL (53). To compare genome-wide ANI scores
246 of Accumulibacter references with pairwise nucleotide identity of the *ppk1* locus of all
247 references, we performed pairwise BLAST for all *ppk1* coding regions and reported the
248 percent identity (54).

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

250 We identified core and accessory gene content of Accumulibacter clades through
251 clustering orthologous groups of genes (COGs) using PyParanoid (55). We used the four
252 Accumulibacter MAGs generated in this study (Table 1) and high-quality Accumulibacter

253 reference genomes (above 90% completeness and less than 5% redundancy) for this
254 analysis, with *Dechloromonas aromatica RCB* and *Rhodocyclus tenuis DSM 110* as
255 outgroups (Accumulibacter references listed in Supplementary Table 1 available at
256 https://figshare.com/articles/dataset/Accumulibacter_References_Metadata/13237148).
257 Briefly, an all-vs-all comparison of proteins for all genomes was performed using
258 DIAMOND (56), pairwise homology scores calculated with the InParanoid algorithm (57),
259 gene families constructed with MCL (58), and HMMs built for each gene family with
260 HMMER (59). A phylogenetic tree was constructed for genomes used for the COGs
261 analysis by using the coding regions of the *ppk1* locus and overlaid with the
262 presence/absence of core and flexible gene content using the ete3 python package (60).
263 A similarity score between UW Accumulibacter genomes was calculated as the pairwise
264 Pearson correlation coefficient between gene families using the numpy python package
265 (61). The presence and absence of different gene families were visualized in an upset
266 plot between UW Accumulibacter genomes using the ComplexUpset package based on
267 the UpSetR package in R (62). All data wrangling and plotting were performed using the
268 tidyverse suite of packages in R (63).

269 **Metatranscriptomics Mapping and Processing**

270 Metatranscriptomics reads from each of the seven samples were quality filtered
271 using fastp (64) and rRNA removed with SortMeRNA (65). The coding regions of all four
272 Accumulibacter genomes assembled in this study (Table 1) were predicted with Prodigal
273 (66) and annotated with both Prokka and KofamKOALA (47, 48), and concatenated
274 together to create a mapping index. Metatranscriptomic reads were competitively
275 pseudoaligned to the reference index and counts quantified with kallisto (67). Genes with

276 a sum of more than 10,000 counts across all 7 samples were removed, as these are likely
277 ribosomal RNAs that were not removed previously. Additionally, genes annotated as
278 rRNAs by barrnap were manually removed (47). Low-count genes were removed by
279 requiring that all genes had to have more than 10 counts mapped across 3 or more
280 samples.

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 (68). Differential 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. We analyzed
287 expression patterns of differentially expressed genes using a threshold cutoff of $\pm$1.5 log2
288 fold-change between anaerobic and aerobic conditions and then identified differentially
289 expressed genes as core or accessory among clades IA (str. UW4) and IIC (str. UW6).
290 For the core group analysis, the corresponding ortholog in each genome had to meet the
291 minimum count threshold as described above to include in the comparison, with genes in
292 one or both of the strains being differentially expressed. For the accessory analysis for
293 each strain, the gene had to be differentially expressed and not contained in the other
294 genome but could be present in other strains or other genomes belonging to that clade.

295 **Data and Code Availability**

296 Raw sequencing files for the three metagenomes and seven metatranscriptomes
297 have been submitted to NCBI under the BioProject accession PRJNA668760. Assembled
298 UW5 and UW6 genomes have been deposited through IMG at accession numbers

299 2767802316 and 2767802455 respectively. All genomes assemblies used in this study in
300 their current forms as well as transcriptomic count tables are available on Figshare at:
301 https://figshare.com/projects/Metabolic_Plasticity_of_Accumulibacter_Clades/90614. All
302 code is available at https://github.com/elizabethmcd/R3R4.
303

304 **RESULTS**

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

306 We inoculated a 2-L sequencing-batch reactor with activated sludge from a full-
307 scale wastewater treatment plant in Madison, WI, USA and maintained operation for 40
308 months. The reactor was primarily fed with acetate and operated under cyclic anaerobic-
309 aerobic phases with a 4-day solids retention time to simulate the EBPR process and
310 enrich for Accumulibacter (Figure 1). During a period of stable EBPR (14 months following
311 inoculation) in which acetate was completely diminished by the start of the aerobic phase
312 and soluble phosphate was below 1.0 mg-L^{-1} at the end of the aerobic phase, we collected
313 seven samples across a single cycle (three in the anaerobic phase, four in the aerobic
314 phase) for RNA-sequencing (Figure 1A). We performed qPCR of the *ppk1* as described
315 previously (29) to quantify Accumulibacter clades present in the reactor on the same day
316 as the RNA-seq experiment (Figure 1B). At the time of the RNA-seq experiment, the
317 bioreactor was primarily enriched in Accumulibacter clade IA, followed by clade IIC and
318 clade IIA.

319 From metagenome samples collected from this bioreactor earlier, we assembled
320 high-quality genomes of belonging to four different Accumulibacter clades (Table 1),
321 including the dominant IA and IIC (strains IA-UW4 and IIC-UW6, respectively) clades and

322 the less abundant IIA and IIF clades (strains IIA-UW5 and IIF-UW7, respectively) (Figure
323 2). Each of the assembled genomes falls within the established Accumulibacter clade
324 nomenclature as defined by *ppk1* sequence identity and phylogenetic placement when
325 compared to other publicly available Accumulibacter genomes and clone sequences
326 (Figure 2A). Additionally, the *ppk1* hierarchical structure is also reflected by pairwise-
327 average nucleotide identity (ANI) boundaries between clades, with some exceptions for
328 newly assembled Accumulibacter genomes that fall outside of the established '*C.*
329 Accumulibacter phosphatis' lineage (Figure 2B). All assembled genomes are above 90%
330 complete and contain less than 5% redundancy as calculated by CheckM (46) (Table 1).
331 We were able to scaffold the assemblies of strains IIA-UW5 and IIC-UW6 using long
332 Nanopore reads, constructing a completely contiguous draft genome of the IIC-UW6
333 strain. To our knowledge, the assembled IIC-UW6 genome is the most contiguous,
334 highest-quality reference genome available for this clade, providing a valuable new
335 Accumulibacter reference genome (Figure 2C).

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

337 We characterized the functional diversity of different Accumulibacter clades
338 through clustering orthologous groups of genes (COGs) (Figure 3A). We repeated a
339 COGs analysis from Skennerton et al. (23) and Oyserman et al. (69), but included an
340 updated set of high-quality Accumulibacter references, the UW-generated genomes, and
341 the outgroups *Dechloromonas aromatica* and *Rhodocyclus tenuis* (Figure 3). The
342 Accumulibacter lineage exhibits a tremendous amount of functional diversity, with only
343 approximately 25% of COGs shared amongst all Accumulibacter clades (Figure 3A).
344 Additionally, there is extensive diversity within the two Accumulibacter types and within

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

352 We then compared COGs between Accumulibacter clades collected from
353 bioreactors seeded from the Nine Springs Wastewater Treatment Plant in Madison, as
354 these genomes represent a shared origin (Figure 3B). In addition to two lab-scale
355 bioreactors harboring distinct Accumulibacter clades (strains UW1 and UW3, and UW4-
356 7, respectively) (22, 30), we recently characterized clade IC (strain UW-LDO-IC) using
357 genome-resolved metagenomics and metatranscriptomics (14). Overall, genomes in
358 Type I and Type II cluster respectively, where genomes within Type I are overall more
359 similar to each other, as expected (Supplementary Figure 1). Clade IIA genomes (strains
360 UW1 and UW5) were assembled from separate enrichments derived from the same full-
361 scale treatment plant approximately 12 years apart, but are very similar by COG
362 presence/absence (Supplementary Figure 1). UW genomes within Type II are less similar
363 by gene content and seem to be more diverse in this regard, though again this could be
364 due to under-sampling of Type I (8 genomes in Type I and 14 in Type II). Additionally,
365 clade IIF contains more overlap in gene content with Type I genomes in gene families
366 than some Type II genomes do to each other (Supplementary Figure 1).

367 COG overlap and uniqueness is shown by the intersection of COG groups between
368 UW Accumulibacter clades and the *Rhodocyclus* and *Decholoromonas* outgroups (Figure
369 3B). There are 1075 gene families shared among all UW Accumulibacter clades and
370 outgroup genomes, with 277 separate gene families only within the UW Accumulibacter
371 clades (Figure 3B). Among individual clades, there are 887 groups only within clade IIC
372 (UW6), 399 within clade IIA (UW1 and UW5), 399 within IIF (UW7), 292 within clade IA
373 (UW3 and UW4), and 196 within IC (UW-LDO). Although clade IIC str. UW6 contains
374 more overlap in gene content with SK-02 (Figure 3A), the intersection analysis (Figure
375 3B) identified gene families that are available to individual strains within the same
376 bioreactor ecosystem, highlighting the potential for functional differentiation.

377 **Transcriptional Profiles of Accumulibacter Strains**

378 Although *ppk1* locus quantification showed that the bioreactor was mostly enriched
379 in clade IA (see Figure 1B), surprisingly, more RNA-seq reads mapped to the clade IIC-
380 UW6 genome than to the clade IA-UW4 (Figure 4A). Approximately 40 million total
381 transcriptional reads mapped back to the strain IIC-UW6 genome across the full time-
382 series experiment, whereas strains IA-UW4 and IIA-UW5 recruited roughly the same
383 number of reads, with 14 million and 11 million, respectively (Table 2). Markedly lower
384 read levels mapped to the IIF-UW7 strain relative to the other three clades, and therefore
385 we do not include the IIF-UW7 strain in our subsequent analyses. We explored the top
386 differentially expressed genes in strain IIC-UW6, comparing anaerobic and aerobic
387 conditions (Figure 4B). Genes upregulated in the anaerobic phase belonged to pathways
388 for central-carbon transformations, fatty acid biosynthesis, and PHA synthesis.
389 Interestingly, all genes within the high-affinity phosphate transport system Pst were

390 upregulated in the aerobic phase. This includes the substrate-binding component *pstS,*
391 *phoU* regulator, and the ABC-type transporter *pstABC* complex (Figure 4B).
392 The co-existing clades in the bioreactor system differ in both the presence of sets
393 of nitrogen cycling gene sets and their expression profiles (Figure 4C). In strain IA-UW4,
394 we detected the *napAGH* genes for nitrate reduction, whereas we did not detect *napB*,
395 possibly due to these genes being close to the end of a contig in this genome. We also
396 detected a *nirS*-like nitrite reductase*,* both subunits of the nitric oxide reductase *norQD,*
397 and the nitrous oxide reductase *nosZ* in strain IA-UW4. Most of these genes in strain IA-
398 UW4 exhibit the highest expression patterns in the early and late anaerobic stages
399 (Figure 4C). Conversely, strain IIC-UW6 contains the *narGHIJ* machinery for nitrate
400 reduction, a different *nirS*-like nitrite reductase than in IA-UW4, and does not contain a
401 nitrous oxide reductase. However, strain IIC-UW6 does contain the nitrite reductase *nirBD*
402 for reducing nitrite to ammonium, which is uniformly expressed across the cycle (Figure
403 4C).
404 **Expression Dynamics of Genes Shared by Strains IA-UW4 and IIC-UW6**
405 When characterizing expression dynamics of COGs shared between strains IA-
406 UW4 and IIC-UW6, we found very few orthologs that exceeded the 1.5-fold change cutoff
407 for differential expression in both strains (Figure 5). These genes include a
408 coproporphyrinogen III oxidase, NAD-reducing HoxS subunits, the *phaC* subunit of the
409 PHA synthase, the long-chain fatty-acid-CoA ligase FadD13, and several hypothetical
410 proteins. The fatty-acid CoA-ligase incorporates ATP, CoA, and fatty acids of variable
411 length to form acyl-CoA to be degraded for energy production, incorporated into complex
412 lipids, or used in other metabolic pathways (70). The acyl-CoA ligase is upregulated in

413 the anaerobic phase of both clades, suggesting a role in anaerobic carbon metabolism,
414 as acyl-CoA can ultimately form acetyl-CoA through beta-oxidation. The *phaC* subunit of
415 the PHA synthase is differentially expressed in the anaerobic phase for both clades, as
416 is typical of the hallmark EBPR metabolism for Accumulibacter (13, 71).

417 Additionally, both the IA-UW4 and IIC-UW6 strains carry most of the subunits for
418 the bidirectional NiFe Hox system, encoding an anaerobic hydrogenase. The HoxH and
419 HoxY beta and delta subunits form the hydrogenase moiety, and the FeS-containing
420 HoxF and HoxU alpha and gamma subunits catalyze NAD(P)H/NAD(P)+
421 oxidation/reduction coupled to the hydrogenase moiety (72–74). In both clades, the FeS
422 cluster HoxF and HoxU alpha and gamma subunits are strongly differentially expressed
423 in the anaerobic phases, as was observed previously in a genome-resolved
424 metatranscriptomic investigation of strain IIA-UW1 (13). However, the IIC-UW6 genome
425 is missing the HoxY delta subunit of the hydrogenase moiety, and only strain IA-UW4
426 exhibits differential expression of the HoxH beta subunit in the anaerobic phase. Strain
427 IIC-UW6 does exhibit higher expression of HoxH in the late anaerobic and early aerobic
428 phases but does not meet the threshold cutoff for differential gene expression. A prior
429 study focused on strain IIA-UW1 also demonstrated hydrogenase activity in the anaerobic
430 phase, and then demonstrated hydrogen gas production after acetate addition (13). The
431 hydrogen gas production was hypothesized to replenish NAD+ after glycogen
432 degradation. Subsequent comparative genomics analysis suggested that the Hox system
433 is an ancestral state of Accumulibacter (69). Furthermore, recent metabolic models which
434 integrated hydrogen production under periods of excess acetate or glycogen were better
435 able to capture the diversity of observed EBPR stoichiometry (75). In this ecosystem,

436 both strains may exhibit anaerobic hydrogenase activity and thus produce hydrogen, but
437 due to the missing subunit in the IIC-UW6 genome and lack of differential gene
438 expression, this is uncertain.

439 Interestingly, only one strain exhibited differential expression patterns of several
440 core ortholog subsets important for EBPR-related functions exhibit. For example, subunits
441 of the high-affinity phosphate transporter Pst system are only differentially expressed in
442 strain IIC-UW6 and not IA-UW4, as observed for strain IIC-UW6 above (see Figure 4B).
443 Additionally, nitrogen cycling genes that contain orthologs in both clades display different
444 expression profiles. A nitrite reductase is only differentially expressed in the anaerobic
445 phase in strain IA-UW4 and the denitrification protein NirQ is differentially expressed in
446 only strain IIC-UW6. Whereas the nitrogen fixation protein subunit NifU is differentially
447 expressed in both strains. To activate acetate to acetyl-CoA, the high-affinity acetyl-CoA
448 synthase (ACS) or low-affinity acetate kinase/phosphotransacetylase (AckA/Pta)
449 pathways can be employed. The strain IA-UW4 assembly is missing an acetate kinase,
450 and phosphate acetyltransferase is only differentially expressed in this strain but not in
451 IIC-UW6. Within strain IIC-UW6, both the acetate kinase and phosphate acetyltransferase
452 genes exhibit relatively low levels of expression over the time-series. Both strains contain
453 several acetyl-CoA synthases that are highly expressed in the anaerobic phase,
454 particularly upon acetate contact, but do not exhibit notable differences in differential
455 expression between the two phases. Previous work based on community proteomics and
456 enzymatic assays have suggested that although both activation pathways are present
457 and expressed in Accumulibacter, the high-affinity pathway is preferentially used (76, 77).
458 Although a phosphate acetyltransferase is differentially expressed in strain IA-UW4, this

459 gene has lower transcripts relative to the acetyl-CoA synthase, providing more evidence
460 that the lineage (specifically both clades IA and IIC) may employ the high-affinity system.

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

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

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

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

492 **DISCUSSION**

493 In this work, we performed time-series transcriptomics to explore gene expression
494 profiles of Accumulibacter strains belonging to clades IA and IIC, which co-existed within
495 the same bioreactor ecosystem. By applying genome-resolved metagenomic and
496 metatranscriptomic techniques, we were able to postulate niche-differentiating features
497 of distinct Accumulibacter clades based on core and accessory gene content and
498 differential expression profiles. We assembled high-quality genomes from four
499 Accumulibacter clades and used long reads to generate a single contiguous scaffold of
500 the IIC-UW6 genome. By performing an updated clustering of orthologous groups of
501 genes analysis, we were able to better demonstrate the similarity and uniqueness among
502 Accumulibacter strains from this system. Unexpectedly, although clade IA dominated in
503 abundance by *ppk1* locus quantification, clade IIC was more transcriptionally active. By
504 comparing the expression profiles of COGs shared between strains IA-UW4 and IIC-

505 UW6, we found that genes involved in EBPR-related pathways may differ between the
506 strains with respect to transcription-level regulation. Notably, strain IIC-UW6 showed a
507 strong upregulation of the high-affinity Pst phosphorus transport system in the aerobic
508 phase. High activity of the Pst system in IIC-UW6 towards the end of the aerobic phase
509 when phosphorus concentrations were lower combined with contrasting abundance vs.
510 transcriptional activity patterns suggests that the IIC-UW6 strain is specialized under
511 phosphorus-limiting conditions. Whereas lack of differential expression of the Pst system
512 in IA-UW4 suggests this strain is more specialized under high phosphorus conditions.
513 Strain IA-UW4 contains most of the subunits for the individual key steps for denitrification,
514 and exhibits strong anaerobic upregulation of these genes, suggesting a differentiating
515 role for strain IA-UW4 in nitrogen cycling under denitrifying conditions. Additionally, the
516 IA-UW4 strain contains all subunits for encoding the Hox anaerobic hydrogenase system,
517 which exhibits upregulated gene expression in the anaerobic phase. However, the IIC-
518 UW6 strain is missing one of the subunits even though the other subunits were expressed
519 (but not differentially). Production of hydrogen gas could be a differentiating role for the
520 IA-UW4 strain, as this would impact potential interactions with diverse flanking community
521 members in addition to replenishing NAD+ (13, 30, 78).

522 The Accumulibacter lineage harbors extensive diversity between and amongst
523 individual clades which is exhibited not only by *ppk1*-based phylogenies, but also by
524 pairwise genome-wide ANI and COG comparisons. For example, strain IIA-UW1 and
525 strain IA-UW3 were present in the same bioreactor ecosystem and only share 85% ANI
526 (22). In the bioreactor enrichment from this study, the dominant strains IA-UW4 and IIC-
527 UW6 share less than 80% ANI (Figure 2B). Recently, a cutoff of >95% ANI has been

528 suggested to delineate distinct bacterial species or cohesive genetic units (51), which is
529 based on observations of coverage gaps at 90-95% sequence identity when
530 metagenomic reads were mapped back to assembled MAGs (79–82). These sequence
531 cutoffs have also been benchmarked against homologous recombination rates and ratios
532 of nonsynonymous to synonymous (*dN/dS*) nucleotide differences (83). However, there
533 are some notable exceptions to the 95% ANI threshold cutoff used for delineating species,
534 as distinct populations of the marine SAR11 clade have been observed at a 92%
535 boundary and lower (84, 85). Among Accumulibacter genomes adhering to the
536 established *ppk1*-based phylogeny, genomes within individual clades fall within the >95%
537 ANI threshold, whereas ANI comparisons within and between Types I and II are more
538 variable (Figure 2B). Genomes within Type I are similar by approximately 90% ANI,
539 whereas genomes within Type II are not as coherent.

540 The apparent coexistence of diverse Accumulibacter clades within an ecosystem
541 then raises questions of ecological roles, putative interactions, and evolutionary dynamics
542 of this lineage over time and space. Although the concept of a bacterial species is highly
543 contested and dependent upon the operational definition of the population in question,
544 species have been clustered based on phenotypic and metabolic characteristics, and
545 more recently by DNA-based clustering metrics and recombination signatures (3, 7). The
546 ecotype model of speciation suggests that cohesive, ecologically distinct ecotypes can
547 coexist because they occupy separate niche spaces (3, 86). Because Accumulibacter
548 has not been isolated in pure culture to date, most ecological and evolutionary inferences
549 have been made from genome-based comparisons (23, 69) or mapping metagenomic
550 reads back to assembled genomes with minimum percent identity cutoffs that may not

551 sufficiently distinguish among sub-clades (87, 88). However, the apparent diversity within
552 the Accumulibacter lineage may require reevaluation of the phylogeny and taxonomy
553 overall. Although the *ppk1*-based clade designations coincide with genome-wide ANI
554 based cutoffs remarkably well (Supplementary Figure 2B), there are some discrepancies.
555 A few Accumulibacter genomes have been assembled that do not adhere to the *ppk1*-
556 based population structure but fall within the Accumulibacter lineage based on a species
557 tree of single-copy core genes (Supplementary Figure 2A), such as the proposed novel
558 species '*Candidatus* Accumulibacter aalborgensis' (89). Additionally, a few genomes
559 preliminarily classified by the Genome Taxonomy Database (GTDB) as Accumulibacter
560 branch outside of the established nomenclature and are quite divergent based on the
561 species and *ppk1* phylogenies as well as genome-wide and *ppk1* sequence similarities
562 (Figure 2B, Supplementary Figure 2). Overall, substantial work is needed to reevaluate
563 the phylogenetic structure of the Accumulibacter lineage as a whole and to connect with
564 signatures of homologous recombination and evolutionary trajectories within coexisting
565 clades to understand the genetic boundaries and thus ecological interdependencies of
566 individual clades or strains.

567 Although the boundaries of individual Accumulibacter clades or strains have not
568 been entirely resolved to infer separate ecotypes, denitrification capabilities of specific
569 clades have been repeatedly suggested to be a niche-differentiating feature (14, 24, 29).
570 Previously, two separate bioreactors fed with either acetate or propionate as the sole
571 carbon source were enriched with different Accumulibacter morphotypes, which were also
572 linked to different denitrification capabilities (25). A study with a bioreactor highly enriched
573 in clade IC inferred simultaneous use of oxygen, nitrite, and nitrate as electron acceptors

574 under micro-aerobic conditions either due to metabolic flexibility or heterogeneity within
575 IC members (29). The resident strain IC-UW-LDO also possessed the full suite of genes
576 for denitrification and expressed them in microaerobic conditions (14). Coexisting
577 Accumulibacter strains with differing denitrifying abilities have also been observed in a
578 denitrifying EBPR bioreactor enriched in clade IA and clade IC (17, 28). The clade IA
579 strain encoded genes for complete denitrification, whereas the co-occurring clade IC
580 strain and other flanking community members did not (28). Additionally, Gao et al.
581 performed an ancestral genome reconstruction analysis similar to Oyserman et al. 2017
582 (69) to specifically understand the flux of denitrification gene families (28). The *nir* gene
583 cluster was identified as a core COG to all sampled Accumulibacter genomes, whereas
584 the periplasmic nitrate reductase *napAGH* and nitrous oxide reductase *nosZDFL* clusters
585 were core to only Type I Accumulibacter genomes, and inferred to have been lost for
586 some Type II genomes (28). Additionally, Wang et al. subsequently observed that an
587 internal stop codon within the open reading frame of the *nosZ* gene in a clade IC strain
588 coincided with decreased transcripts mapping back (17, 28). In our study, we observed
589 differential expression of denitrification genes in strain IA-UW4 but not strain IIC-UW6.
590 These expression differences could explain why the two strains can coexist and
591 contribute to overall ecosystem functioning (e.g. phosphorus removal) by using different
592 electron acceptors in the anaerobic/aerobic conditions. Although the synthetic feed for
593 the bioreactors in this study inhibits ammonia oxidation through allylthiourea addition (and
594 thus inhibits nitrite/nitrate production to fuel respiratory denitrification), the expression
595 profiles of denitrification genes in strain IA-UW4 suggests this strain is equipped to have
596 a differentiating role in a denitrifying bioreactor. We specifically observed differential

597 expression of nitrate reductase and nitrous-oxide reductase genes in strain IA-UW4,
598 where homologs of these genes were not present in the strain IIC-UW6 genome and other
599 Type II genomes. From our work and previous studies, lack of the complete genetic
600 repertoire for denitrification among and within individual clades in addition to
601 transcriptional-based evidence suggests that denitrification capabilities could be a niche
602 differentiating role for specific Accumulibacter populations.

603 Accumulibacter's anaerobic metabolism has also been shown to differ under
604 varying substrate or stoichiometric conditions and possibly between the two
605 Accumulibacter types. Recent simulations of 'omics datasets paired with enzymatic
606 assays hint that the routes of anaerobic metabolism for Accumulibacter may differ based
607 on environmental conditions (75). Under high acetate conditions, polyphosphate
608 accumulation arises through the glyoxylate shunt, whereas under increased glycogen
609 concentrations, glycolysis is used in conjunction with the reductive branch of the TCA
610 cycle (75). Furthermore, Welles et al. found that Type II Accumulibacter can switch to
611 partial glycogen degradation more quickly than Type I, enabling Type II to fuel VFA uptake
612 in polyphosphate-limited conditions (90, 91). Conversely, when concentrations of
613 polyphosphate are high, VFA uptake rates are higher in Type I. However, these studies
614 only identified Accumulibacter lineages based on *ppk1* types, and therefore extrapolated
615 metabolic assumptions from presumably more resolved strains to the entire type. These
616 broad interpretations may obscure apparent functional differences between the two types
617 and individual clades. Future experiments could integrate multi-'omics approaches to
618 investigate not only metabolic flexibility under different substrate conditions and
619 perturbations but also how metabolisms are distributed amongst Accumulibacter clades

620 in bioreactor ecosystems that are not completely enriched in a single clade or strain and
621 elucidate key functional differences.
622
623 **CONCLUSIONS**
624 In this study, we integrated genome-resolved metagenomics and time-series
625 metatranscriptomics to understand gene expression patterns of co-existing
626 Accumulibacter strains within a bioreactor ecosystem. We found evidence for
627 denitrification gene expression in the dominating IA-UW4 strain, but higher overall
628 transcriptional activity for the IIC-UW6 strain along with differential expression of a high-
629 affinity phosphorus transporter. The coexistence of Accumulibacter populations exhibiting
630 different gene expression patterns for phosphorus and nitrogen cycling and hydrogen
631 productions suggest niche differentiating features and contribution to different metabolic
632 programs. Detailed experiments incorporating fluorescent in situ hybridization combined
633 with microautoradiography could be applied to confirm metabolic activities of coexisting
634 Accumulibacter populations and how they contribute to important EBPR-related functions
635 (92). These results also highlight the need to reevaluate the phylogenetic structure of the
636 Accumulibacter lineage and define how potential genetic boundaries may translate to
637 individual ecological niches that allow for the coexistence of multiple clades or strains.
638 Additionally, we highlight an approach for exploring functions of closely related
639 uncultivated microbial lineages by exploring gene expression patterns of shared and
640 accessory genes. Overall, this work emphasizes the need to understand the basic
641 ecology and evolution of microorganisms underpinning biotechnological processes,

642 which can lead to better treatment process control and resource recovery outcomes in
643 the future (93).
644
645
646 **ACKNOWLEDGEMENTS**

647 We thank Dr. Jeffrey Lewis for feedback on the differential expression analyses
648 and interpretation. This work was supported by funding from the National Science
649 Foundation (MCB-1518130) to K.D.M. and D.R.N. F.M. and P.C. received funding from
650 the Chilean National Commission for Scientific and Technological Research (CONICYT).
651 E.H.K. was supported by the 2017 UW-Madison Hilldale Scholarship. E.A.M. was
652 supported by a University of Wisconsin – Madison Department of Bacteriology
653 Fellowship. We thank the University of Wisconsin – Madison Biotechnology Center DNA
654 Sequencing Facility for Illumina sequencing services. This research was performed in
655 part using the Wisconsin Energy Institute computing cluster, which is supported by the
656 Great Lakes Bioenergy Research Center as part of the U.S. Department of Energy Office
657 of Science (DE-SC0018409).
658
659
660
661
662
663
664

REFERENCES

1. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phytogeny of coexisting Prochlorococcus ecotypes. Nature 393:464–467.
2. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R, Chisholm SW. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science (80-) 344:416–420.
3. Cohan FM. 2001. Bacterial Species and Speciation. Syst Biol 50:513–524.
4. N K, SE R, JW B-T, M G, R S, SW C. 2018. Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus.
5. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. 2006. Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science (80-) 311:1737–1740.
6. Voigt K, Sharma CM, Mitschke J, Joke Lambrecht S, Voß B, Hess WR, Steglich C. 2014. Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. ISME J 8:2056–2068.
7. Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: Making sense of genetic and ecological diversity. Science (80-). Science.
8. Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. American Society for Microbiology.
9. Comte J, Fauteux L, Giorgio PA del. 2013. Links between metabolic plasticity and

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

711 phosphorus removal under tropical conditions. Sci Rep 6:1–15.
712 17. Wang Y, Gao H, Wells G. 2020. Integrated Omics Analyses Reveal Differential
713 Gene Expression and Potential for Cooperation Between Denitrifying
714 Polyphosphate and Glycogen Accumulating Organisms. bioRxiv. bioRxiv.
715 18. Mcmahon KD, Dojka M a, Pace NR, Jenkins D, Keasling JD. 2002.
716 Polyphosphate Kinase from Activated Sludge Performing Enhanced Biological
717 Phosphorus Removal Polyphosphate Kinase from Activated Sludge Performing
718 Enhanced Biological Phosphorus Removal †. Appl Environ Microbiol 68:4971–
719 4978.
720 19. He S, Gall DL, McMahon KD. 2007. “Candidatus accumulibacter” population
721 structure in enhanced biological phosphorus removal sludges as revealed by
722 polyphosphate kinase genes. Appl Environ Microbiol 73:5865–5874.
723 20. Peterson SB, Warnecke F, Madejska J, McMahon KD, Hugenholtz P. 2008.
724 Environmental distribution and population biology of Candidatus Accumulibacter,
725 a primary agent of biological phosphorus removal. Environ Microbiol 10:2692–
726 2703.
727 21. McMahon KD, Yilmaz S, He S, Gall DL, Jenkins D, Keasling JD. 2007.
728 Polyphosphate kinase genes from full-scale activated sludge plants. Appl
729 Microbiol Biotechnol 77:167–173.
730 22. Flowers JJ, He S, Malfatti S, del Rio TG, Tringe SG, Hugenholtz P, McMahon KD.
731 2013. Comparative genomics of two “Candidatus Accumulibacter” clades
732 performing biological phosphorus removal. ISME J 7:2301–2314.
733 23. Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. 2015. Expanding our

734 view of genomic diversity in *C andidatus* Accumulibacter clades. Environ
735 Microbiol 17:1574–1585.
736 24. Flowers JJ, He S, Yilmaz S, Noguera DR, McMahon KD. 2009. Denitrification
737 capabilities of two biological phosphorus removal sludges dominated by different
738 "Candidatus Accumulibacter" clades. Environ Microbiol Rep 1:583–588.
739 25. Carvalho G, Lemos PC, Oehmen A, Reis MAM. 2007. Denitrifying phosphorus
740 removal: Linking the process performance with the microbial community structure.
741 Water Res 41:4383–4396.
742 26. Saad SA, Welles L, Abbas B, Lopez-Vazquez CM, van Loosdrecht MCM,
743 Brdjanovic D. 2016. Denitrification of nitrate and nitrite by "Candidatus
744 Accumulibacter phosphatis" clade IC. Water Res 105:97–109.
745 27. Rubio-Rincón FJ, Weissbrodt DG, Lopez-Vazquez CM, Welles L, Abbas B,
746 Albertsen M, Nielsen PH, van Loosdrecht MCM, Brdjanovic D. 2019. "Candidatus
747 Accumulibacter delftensis": A clade IC novel polyphosphate-accumulating
748 organism without denitrifying activity on nitrate. Water Res 161:136–151.
749 28. Gao H, Mao Y, Zhao X, Liu WT, Zhang T, Wells G. 2019. Genome-centric
750 metagenomics resolves microbial diversity and prevalent truncated denitrification
751 pathways in a denitrifying PAO-enriched bioprocess. Water Res 155:275–287.
752 29. Camejo PY, Owen BR, Martirano J, Ma J, Kapoor V, Santo Domingo J, McMahon
753 KD, Noguera DR. 2016. Candidatus Accumulibacter phosphatis clades enriched
754 under cyclic anaerobic and microaerobic conditions simultaneously use different
755 electron acceptors. Water Res 102:125–137.
756 30. Martín HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C,

He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol 24:1263–1269.

31. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.

32. Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

33. Li H. 2013. Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences.

34. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829.

35. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res 40.

36. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26.

37. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–51.

38. Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl JL, Ivanova N, Woyke T, Kyrpides N, Pati A. 2016. ProDeGe: A computational

780 protocol for fully automated decontamination of genomes. ISME J 10:269–272.
781 39. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.
782 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data.
783 PeerJ 3:e1319.
784 40. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I.
785 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long
786 reads. Gigascience 4:35.
787 41. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T, Sun
788 N, Rao J, Wang J, Zhang W, Chen Y, Liao S, Jiang H, Liu X, Yang Z, Mu F, Gao
789 S. 2017. LR_Gapcloser: a tiling path-based gap closer that uses long reads to
790 complete genome assembly. Gigascience 6:1.
791 42. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
792 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler
793 G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly
794 algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–77.
795 43. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read
796 merging via overlap. PLoS One 12:e0185056.
797 44. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for
798 accurately reconstructing single genomes from complex microbial communities.
799 PeerJ 3:e1165.
800 45. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: A tool for fast and
801 accurate genomic comparisons that enables improved genome recovery from
802 metagenomes through de-replication. ISME J 11:2864–2868.

803 46. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
804 assessing the quality of microbial genomes recovered from isolates, single cells,
805 and metagenomes. Genome Res 25:1043–55.
806 47. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
807 30:2068–2069.
808 48. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H.
809 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and
810 adaptive score threshold. bioRxiv 602110.
811 49. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
812 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
813 30:3059–66.
814 50. Stamatakis A. 2014. The RAxML v8 . 0 . X Manual 1–55.
815 51. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
816 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
817 boundaries. Nat Commun 9:5114.
818 52. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to
819 classify genomes with the Genome Taxonomy Database. Bioinformatics.
820 53. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the
821 display and annotation of phylogenetic and other trees. Nucleic Acids Res
822 44:W242–W245.
823 54. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden
824 TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
825 55. Melnyk RA, Hossain SS, Haney CH. 2019. Convergent gain and loss of genomic

826 islands drive lifestyle changes in plant-associated Pseudomonas. ISME J
827 13:1575–1588.
828 56. Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using
829 DIAMOND. Nat Methods. Nature Publishing Group.
830 57. Sonnhammer ELL, Östlund G. 2015. InParanoid 8: Orthology analysis between
831 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43:D234–D239.
832 58. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-
833 scale detection of protein families. Nucleic Acids Res 30:1575–84.
834 59. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol
835 7:e1002195.
836 60. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and
837 Visualization of Phylogenomic Data. Mol Biol Evol 33:1635–1638.
838 61. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D,
839 Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson
840 J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ,
841 Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R,
842 Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F,
843 van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner
844 A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C,
845 Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V., Olivetti E, Martin E,
846 Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen
847 GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J,
848 Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV,

849 Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K,
850 Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ,
851 Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT,
852 Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T,
853 Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T,
854 Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. 2020. SciPy 1.0:
855 fundamental algorithms for scientific computing in Python. Nat Methods 17:261–
856 272.
857 62. Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the
858 visualization of intersecting sets and their properties. Bioinformatics 33:2938–
859 2940.
860 63. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund
861 G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K,
862 Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo
863 K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.
864 64. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
865 preprocessor. Bioinformatics 34:i884–i890.
866 65. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of
867 ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.
868 66. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010.
869 Prodigal: prokaryotic gene recognition and translation initiation site identification.
870 BMC Bioinformatics 11:119.
871 67. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-

872 seq quantification. Nat Biotechnol 34:525–527.
873 68. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
874 dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.
875 69. Oyserman BO, Moya F, Lawson CE, Garcia AL, Vogt M, Heffernen M, Noguera
876 DR, McMahon KD. 2016. Ancestral genome reconstruction identifies the
877 evolutionary basis for trait acquisition in polyphosphate accumulating bacteria.
878 ISME J 10:2931–2945.
879 70. Watkins PA. 2018. The Fatty Acyl-CoA SynthetasesReference Module in
880 Biomedical Sciences. Elsevier.
881 71. He S, McMahon KD. 2011. “Candidatus Accumulibacter” gene expression in
882 response to dynamic EBPR conditions. ISME J 5:329–340.
883 72. Long M, Liu J, Chen Z, Bleijlevens B, Roseboom W, Albracht SPJ. 2007.
884 Characterization of a HoxEFUYH type of [NiFe] hydrogenase from
885 Allochromatium vinosum and some EPR and IR properties of the hydrogenase
886 module. J Biol Inorg Chem 12:62–78.
887 73. Antal TK, Oliveira P, Lindblad P. 2006. The bidirectional hydrogenase in the
888 cyanobacterium Synechocystis sp. strain PCC 6803. Int J Hydrogen Energy
889 31:1439–1444.
890 74. Vignais PM, Billoud B. 2007. Occurrence, classification, and biological function of
891 hydrogenases: An overview. Chem Rev. Chem Rev.
892 75. Guedes da Silva L, Olavarria Gamez K, Castro Gomes J, Akkermans K, Welles L,
893 Abbas B, van Loosdrecht MCM, Wahl SA. 2020. Revealing the metabolic
894 flexibility of Candidatus Accumulibacter phosphatis through redox cofactor

895 analysis and metabolic network modeling. Appl Environ Microbiol.
896 76. Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, Hettich RL,
897 Bond PL, VerBerkmoes NC, Banfield JF. 2008. Community proteogenomics
898 highlights microbial strain-variant protein expression within activated sludge
899 performing enhanced biological phosphorus removal. ISME J 2:853–864.
900 77. Hesselmann RPX, Von Rummell R, Resnick SM, Hany R, Zehnder AJB. 2000.
901 Anaerobic metabolism of bacteria performing enhanced biological phosphate
902 removal. Water Res 34:3487–3494.
903 78. Scarborough MJ, Lawson CE, Hamilton JJ, Donohue TJ, Noguera DR. 2018.
904 Metatranscriptomic and Thermodynamic Insights into Medium-Chain Fatty Acid
905 Production Using an Anaerobic Microbiome. mSystems 3:e00221-18.
906 79. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J,
907 Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG,
908 Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR.
909 2016. Genome-wide selective sweeps and gene-specific sweeps in natural
910 bacterial populations. ISME J 10:1589–1601.
911 80. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T,
912 Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018.
913 Contrasting patterns of genome-level diversity across distinct co-occurring
914 bacterial populations. ISME J 12:742–755.
915 81. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist,
916 metagenomics reveal. Environ Microbiol 14:347–355.
917 82. Konstantinidis KT, DeLong EF. 2008. Genomic patterns of recombination, clonal

918 divergence and environment in marine microbial populations. ISME J 2:1052–
919 1065.
920 83. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield
921 JF. 2020. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial
922 Species Boundaries. mSystems 5.
923 84. Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, Ranjan P,
924 Sarode N, Malmstrom RR, Padilla CC, Stone BK, Bristow LA, Larsen M, Glass
925 JB, Thamdrup B, Woyke T, Konstantinidis KT, Stewart FJ. 2016. SAR11 bacteria
926 linked to ocean anoxia and nitrogen loss. Nature 536:179–183.
927 85. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, Giovannoni S, Eren
928 AM. 2019. Single-amino acid variants reveal evolutionary processes that shape
929 the biogeography of a global SAR11 subclade. Elife 8.
930 86. Cohan FM. 2006. Towards a conceptual and operational union of bacterial
931 systematics, ecology, and evolution, p. 1985–1996. *In* Philosophical Transactions
932 of the Royal Society B: Biological Sciences. Royal Society.
933 87. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, Cordero
934 OX. 2018. Strain-level diversity drives alternative community types in millimetre-
935 scale granular biofilms. Nat Microbiol 3:1295–1303.
936 88. Albertsen M, Benedicte L, Hansen S, Saunders AM, Nielsen H, Lehmann Nielsen
937 K. 2011. A metagenome of a full-scale microbial community carrying out
938 enhanced biological phosphorus removal. ISME J 6:1094–1106.
939 89. Albertsen M, McIlroy SJ, Stokholm-Bjerregaard M, Karst SM, Nielsen PH. 2016.
940 “Candidatus Propionivibrio aalborgensis”: A novel glycogen accumulating

941 organism abundant in full-scale enhanced biological phosphorus removal plants.
942 Front Microbiol 7:1033.
943 90. Welles L, Abbas B, Sorokin DY, Lopez-Vazquez CM, Hooijmans CM, van
944 Loosdrecht MCM, Brdjanovic D. 2017. Metabolic Response of “Candidatus
945 Accumulibacter Phosphatis” Clade II C to Changes in Influent P/C Ratio. Front
946 Microbiol 7:2121.
947 91. Welles L, Tian WD, Saad S, Abbas B, Lopez-Vazquez CM, Hooijmans CM, van
948 Loosdrecht MCM, Brdjanovic D. 2015. Accumulibacter clades Type I and II
949 performing kinetically different glycogen-accumulating organisms metabolisms for
950 anaerobic substrate uptake. Water Res 83:354–366.
951 92. Kim JM, Lee HJ, Lee DS, Jeon CO. 2013. Characterization of the denitrification-
952 associated phosphorus uptake properties of “Candidatusaccumulibacter
953 phosphatis” clades in sludge subjected to enhanced biological phosphorus
954 removal. Appl Environ Microbiol 79:1969–1979.
955 93. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE,
956 O’Malley MA, García Martín H, Pfleger BF, Raskin L, Venturelli OS, Weissbrodt
957 DG, Noguera DR, McMahon KD. 2019. Common principles and best practices for
958 engineering microbiomes. Nat Rev Microbiol. Nature Publishing Group.
959
960
961
962
963

964 **FIGURE AND TABLE LEGENDS**

965

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

967 **1A.** Overview of EBPR cycle and sampling scheme. Samples were collected for RNA
968 sequencing over the course of a normal EBPR cycle, with reported measurements for
969 soluble phosphorus and acetate normalized by volatile suspended solids (VSS). Samples
970 were taken at time-points 0, 31, 70, 115, 150, 190, and 250 minutes. **1B.** Abundance of
971 different Accumulibacter clades by *ppk1* copies per ng of DNA. Quantification of
972 Accumulibacter clades was performed by qPCR of the *ppk1* gene as described in Camejo
973 et al. (29).

974

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

976 **2A.** Phylogenetic tree of *ppk1* nucleotide sequences from available Accumulibacter
977 reference genomes, clone sequences, and the four assembled Accumulibacter genomes
978 from this study. Tips in bold represent *ppk1* sequences from metagenome-assembled
979 genomes, whereas others are sequences from clone fragments available at the
980 referenced accession numbers. Starred tips represent *ppk1* sequences from the four
981 assembled Accumulibacter genomes from this study. Tree was constructed using RaXML
982 with 100 rapid bootstraps. **2B.** Pairwise genome-wide ANI comparisons of all publicly
983 available Accumulibacter references and the four new Accumulibacter genomes, denoted
984 in bold. Genomes are grouped according to clade designation by *ppk1* sequence identity
985 and in order of the *ppk1* phylogeny, except for those that do not fall in an established
986 clade. **2C.** Genome diagram of the UW6 Accumulibacter IIC MAG (98.6% completeness,

987 2.5% contamination, 5.18 Mbp). Layers represent from top to bottom: coverage over 1000
988 bp windows, GC content and skew across the same 1000 bp windows, coding sequences
989 on the positive and negative strands, and positions of predicted rRNA (red) and tRNA
990 (black) sequences.
991
992 **Figure 3. Core and Flexible Gene Content among the Accumulibacter lineage and**
993 **UW genomes**
994 **3A.** Phylogenetic tree of the coding regions for the *ppk1* locus among high-quality
995 Accumulibacter references, UW genomes, and outgroup taxa *Dechloromonas aromatica*
996 *RCB* and *Rhodocyclus tenuis DSM 110.* The pie graph at each node represents the
997 fraction of genes that are shared and flexible for genomes belonging to that clade. **3B.**
998 Upset plot representing the intersections of groups of orthologous gene clusters among
999 UW genomes and outgroups, analogous to a Venn diagram. 1075 gene families were
1000 shared among all UW Accumulibacter clades and outgroup genomes and 277 separate
1001 gene families were found only within the UW Accumulibacter clades. Each bar represents
1002 the number of orthologous gene clusters, and the dot plot represents the genomes in
1003 which the groups intersect.
1004
1005 **Figure 4. Transcriptomic Profiles of Accumulibacter Clades**
1006 **4A.** Average expression profiles of each reference genome in the anaerobic and aerobic
1007 phases. The IIC-UW6 genome mapped the most transcripts even though the IA-UW4 was
1008 more abundant. Reads were competitively mapped to the four assembled Accumulibacter
1009 clade genomes and normalized by transcripts per million (TPM). The sum of the

1010 normalized counts of the three samples in the anaerobic phase and four samples in the
1011 aerobic phase were averaged, respectively, and plotted on a log_{10} scale. Numbers and
1012 percentages for adjacent to normalized bars represent the raw number of reads that are
1013 mapped to each genome. **4B.** Top 50 differentially-expressed genes in clade IIC-UW6.
1014 All 50 genes pass the threshold of ±1.5 log-10 fold-change across the cycle. Annotations
1015 were predicted from a combination of Prokka (47) and KofamKOALA (48). **4C.**
1016 Denitrification gene expression among clades IIC-UW6, IA-UW4, and IIA-UW5. We
1017 summarized expression for genes involved in denitrification by highlighting parts of the
1018 cycle in which certain genes are expressed the most. This is characterized by acetate
1019 contact (pink), early anaerobic (dark purple), late anaerobic (purple), early aerobic (dark
1020 blue) and late aerobic (light blue), where uniform describes uniform expression across
1021 the entire cycle (grey). If an arrow is missing for a step, that particular clade did not contain
1022 confident annotations within the genome for those genes. For nitrate reduction, the *nar*
1023 system is represented as a black dot and the *nap* system is represented as a grey dot.
1024 Multiple arrows are shown for a single step in which particular genes were highly
1025 expressed in multiple phases.
1026
1027 **Figure 5. Differential Expression of Shared Genes in Strains IA-UW4 and IIC-UW6**
1028 Gene expression profiles of core COGs across the EBPR cycle that are differentially
1029 expressed between anaerobic and aerobic conditions in either IA-UW4 or IIC-UW6 or
1030 both strains. To consider a gene differentially expressed in a strain, the COG must exhibit
1031 ±1.5 fold-change in expression between anaerobic and aerobic phases. Homologs that

1032 are directly compared to each other in IA-UW4 and IIC-UW6 are orthologous genes that
1033 belong to the same cluster.
1034
1035 **Figure 6. Differential Expression of Accessory Genes in Strain IA-UW4**
1036 Gene expression profiles of groups of accessory gene groups in strain IA-UW4 that are
1037 not present in the IIC-UW6 genome. A gene was considered differentially expressed
1038 between the anaerobic and aerobic conditions if it exhibited greater or less than ±1.5 fold-
1039 change in expression. Colored dots represent whether or not that gene contains orthologs
1040 within outgroup genomes, other Type II Accumulibacter other than the IIC-UW6 genome,
1041 within Type I Accumulibacter, or only within IA.
1042
1043 **Table 1. Assembled Accumulibacter Genome Statistics**
1044 Genome quality calculations for completeness and contaminations were made with
1045 CheckM based on the presence/absence of single-copy genes (46).
1046
1047 **Table 2. Metatranscriptomic Reads Mapped to Accumullibacter Genomes**
1048 Counts of raw metatranscriptomic reads mapped to each Accumulibacter genome with
1049 kallisto (67) in each sample for the three anaerobic samples and four aerobic samples
1050 and total reads mapped for all samples.
1051
1052 **Supplementary Figure 1.**
1053 Pairwise Pearson correlation coefficient of shared gene content as a measure of how
1054 similar each of the UW genomes is to each other based on COGs.

1055 **Supplementary Figure 2.**

1056 **S2A)** Species tree of all Accumulibacter reference genomes, UW-assembled genomes,
1057 and outgroup genomes using single-copy markers from the GTDB-tk (52). A phylogenetic
1058 tree was constructed using RAxmL with 100 rapid bootstraps and visualized in iTOL. **S2B)**
1059 Pairwise BLAST percent identity scores for *ppk1* coding regions for each reference
1060 genome in Figure 2B, in order of the *ppk1* phylogeny in Figure 2A. Raw pairwise percent
1061 identity scores are available at
1062 https://figshare.com/articles/dataset/Pairwise_Accumulibacter_ppk1_BLAST_Percent_D
1063 _Scores/13237247.

1064

1065 **Supplementary Table 1**

1066 Genome accessions and statistics for all Accumulibacter references used in phylogenies,
1067 ANI comparisons, and COG calculations. A reference genome was used for COG
1068 comparisons if it was a high quality genome (>95% complete and <5% redundant).
1069 Available at https://figshare.com/account/projects/90614/articles/13237148

1070

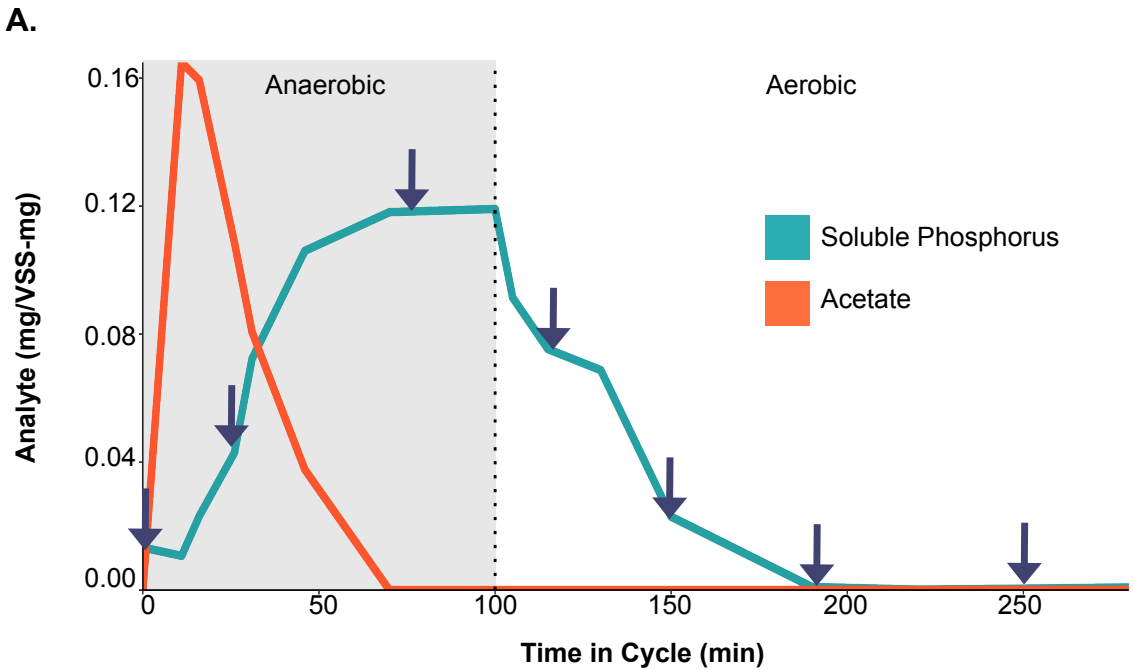
1071 **Supplementary File 1**

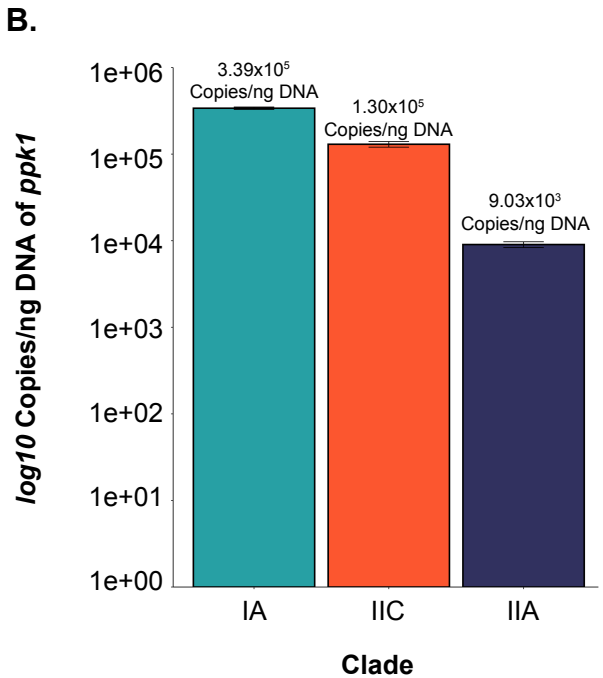
1072 Raw pairwise genome-wide ANI results for each genome calculated with FastANI.
1073 Available at https://figshare.com/account/projects/90614/articles/13237244

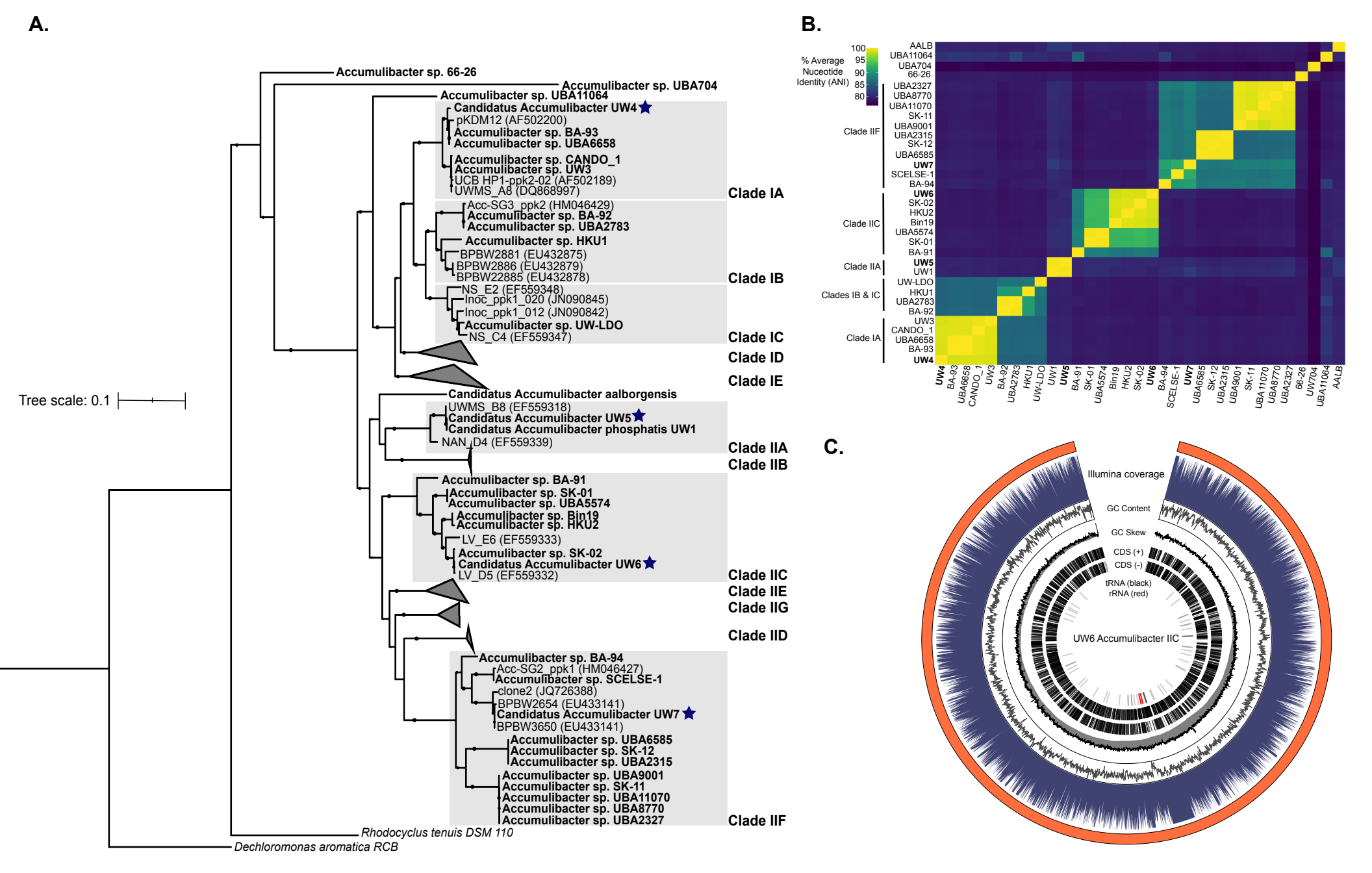
1074

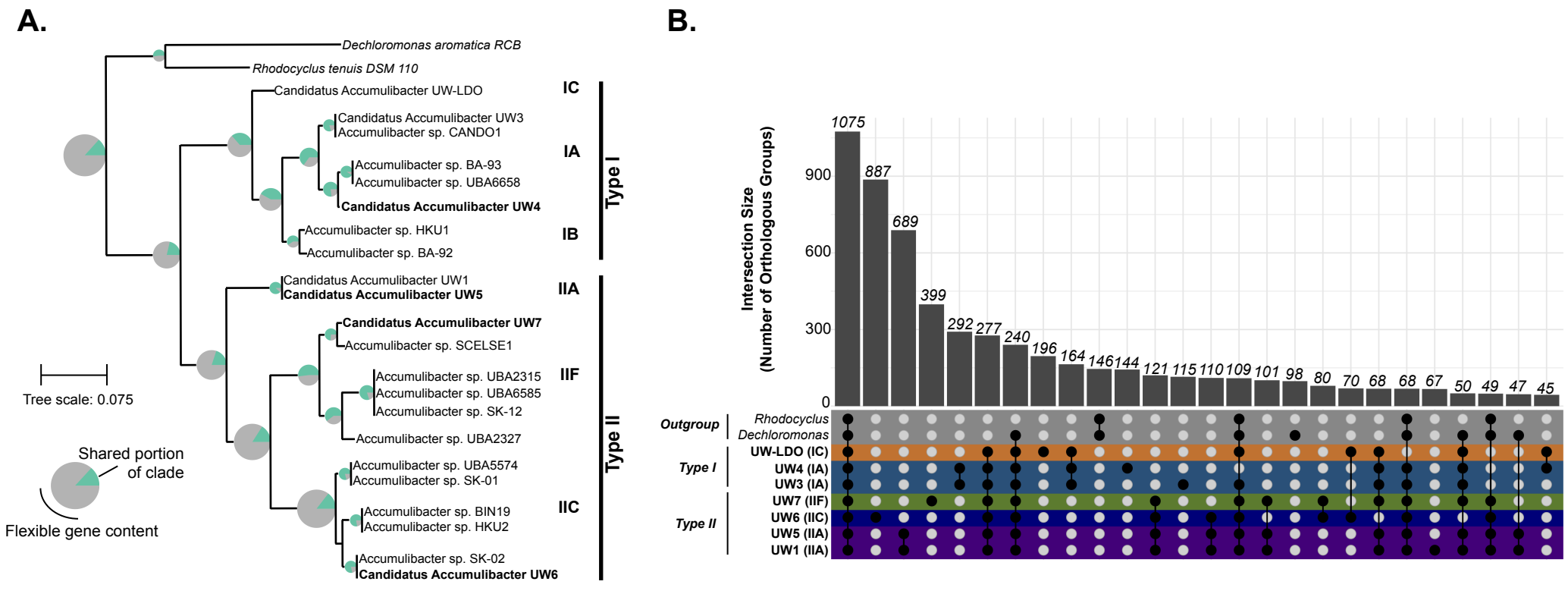
1075 **Supplementary File 2**

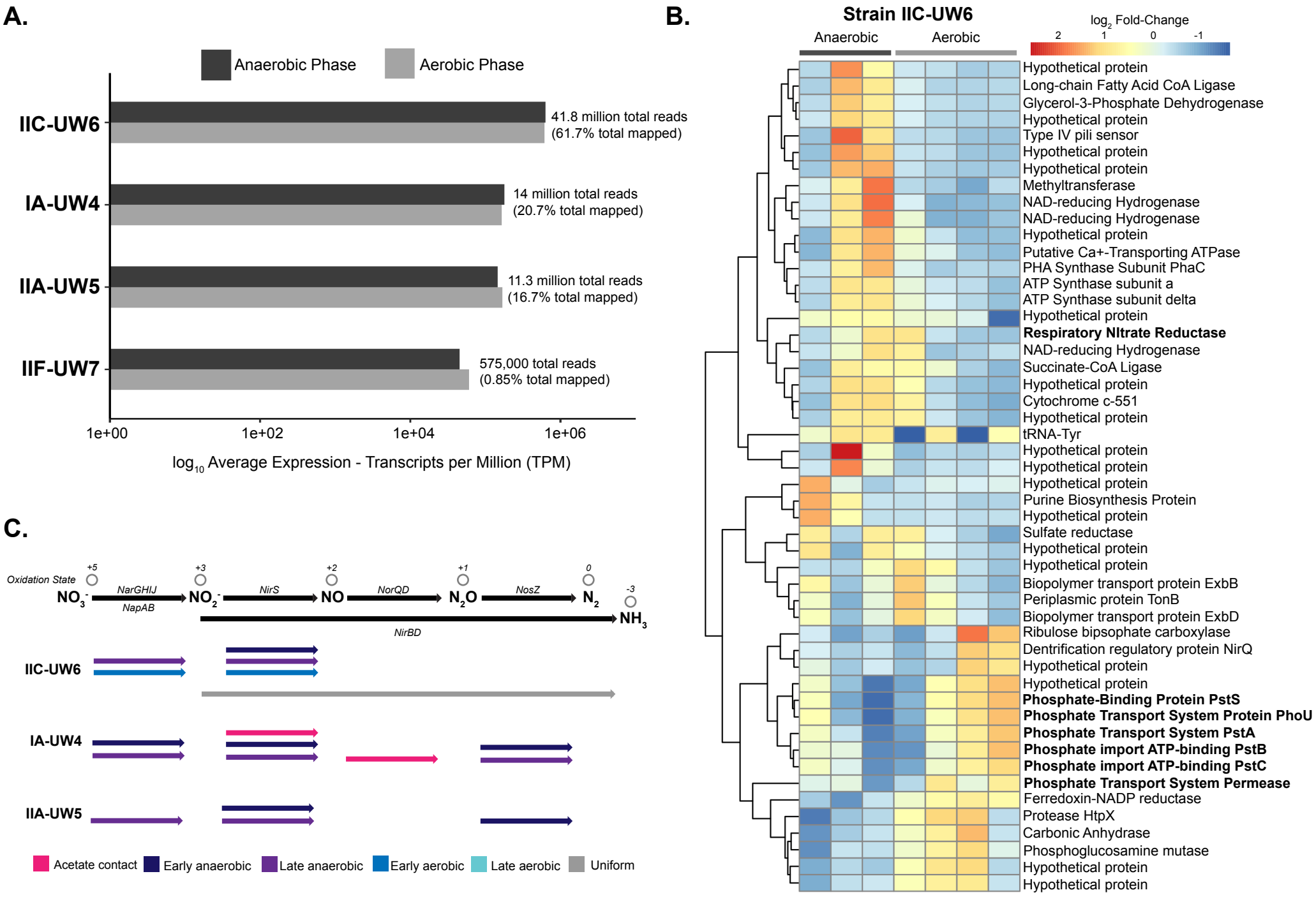
1076 Raw pairwise BLAST percent identity (PID) results for *ppk1* genes from each reference
1077 genome. Available at https://figshare.com/account/projects/90614/articles/13237247

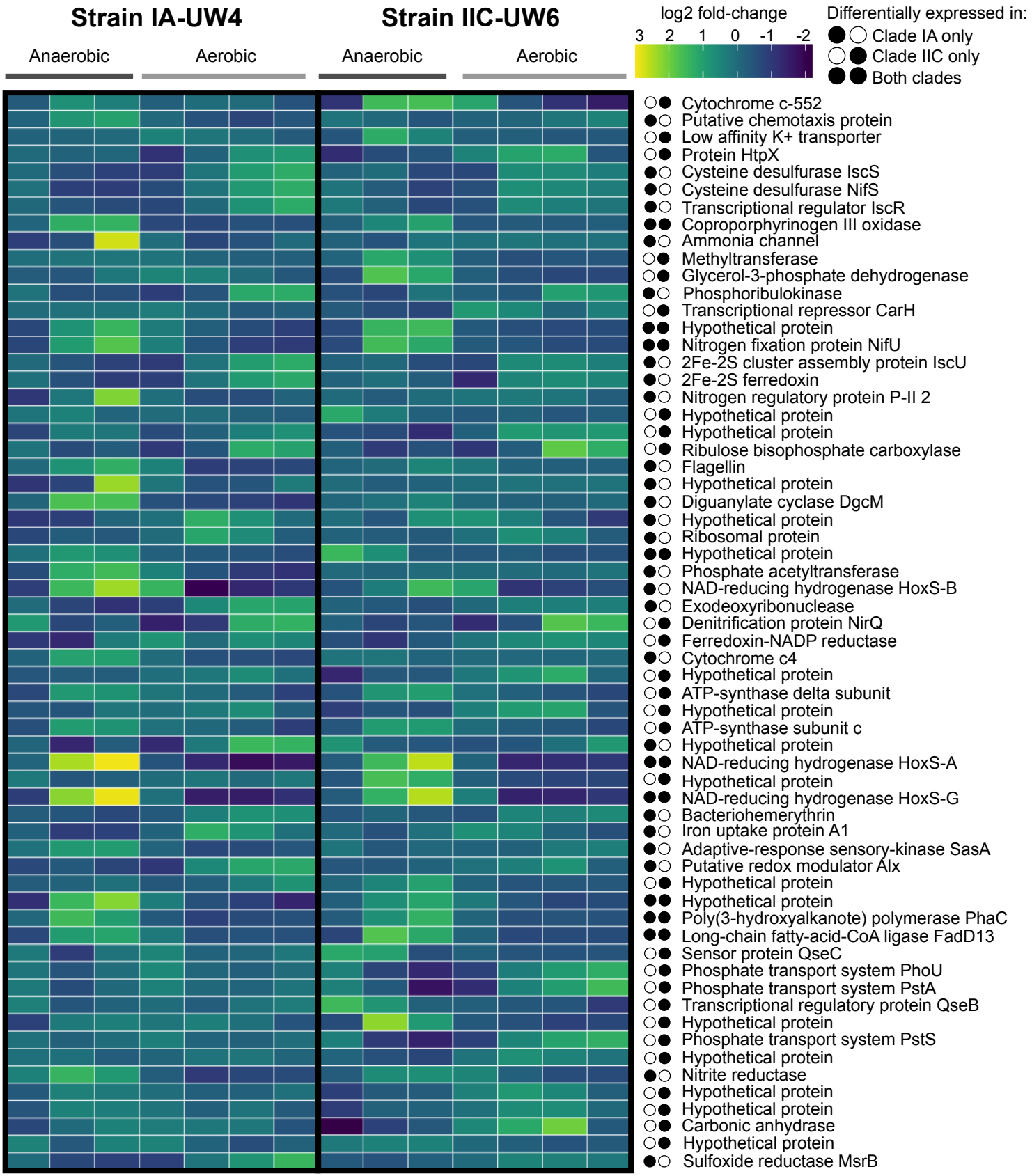

A.
Anaerobic
Aerobic
Soluble Phosphorus
Acetate
Analyte (mg/VSS-mg)
0.16
0.12
0.08
0.04
0.00
0
50
100
150
200
250
Time in Cycle (min)

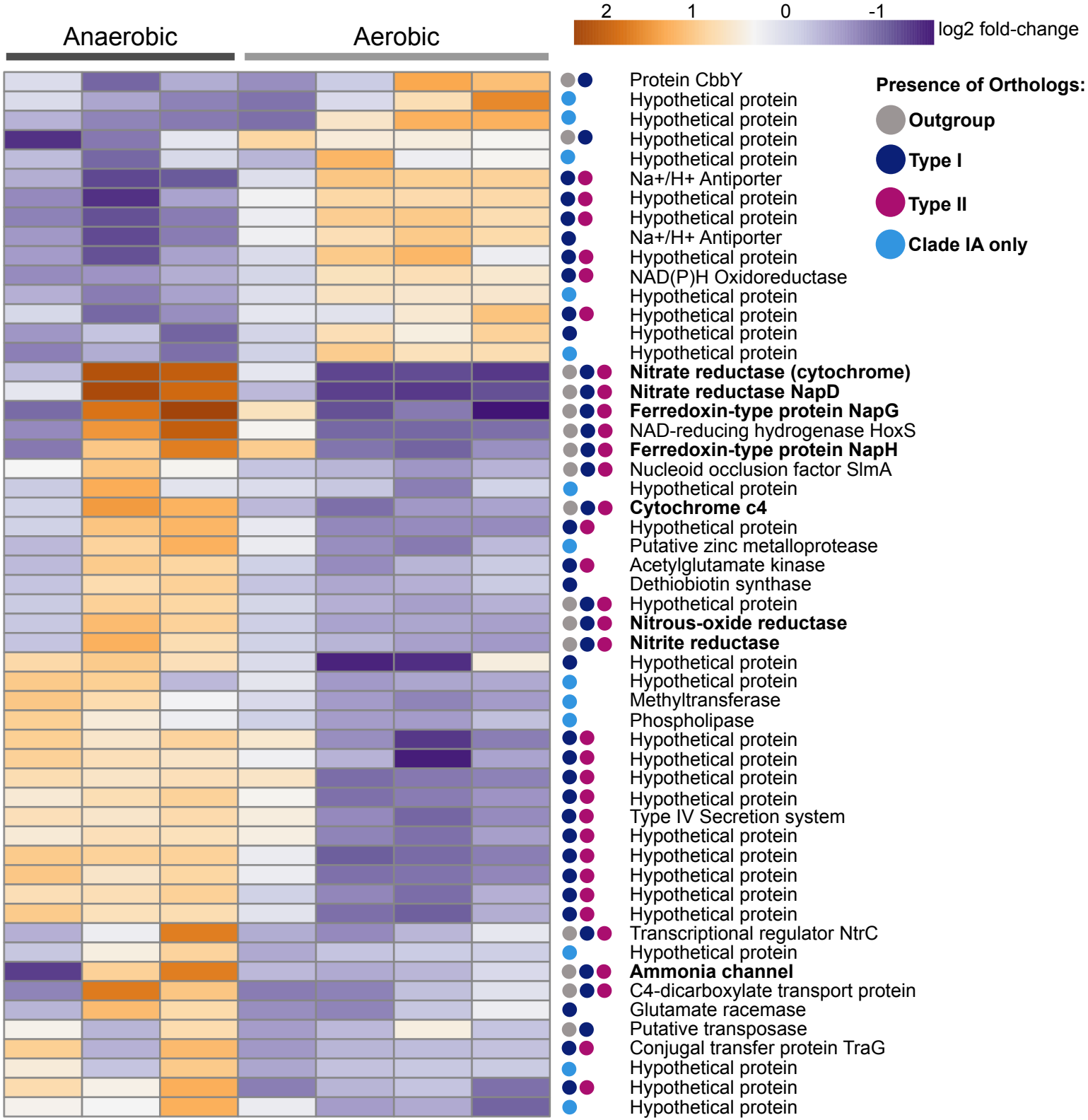

B.
3.39x10^5
Copies/ng DNA
1.30x10^5
Copies/ng DNA
9.03x10^3
Copies/ng DNA
log10 Copies/ng DNA of ppk1
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1e+00
IA
IIC
IIA
Clade


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

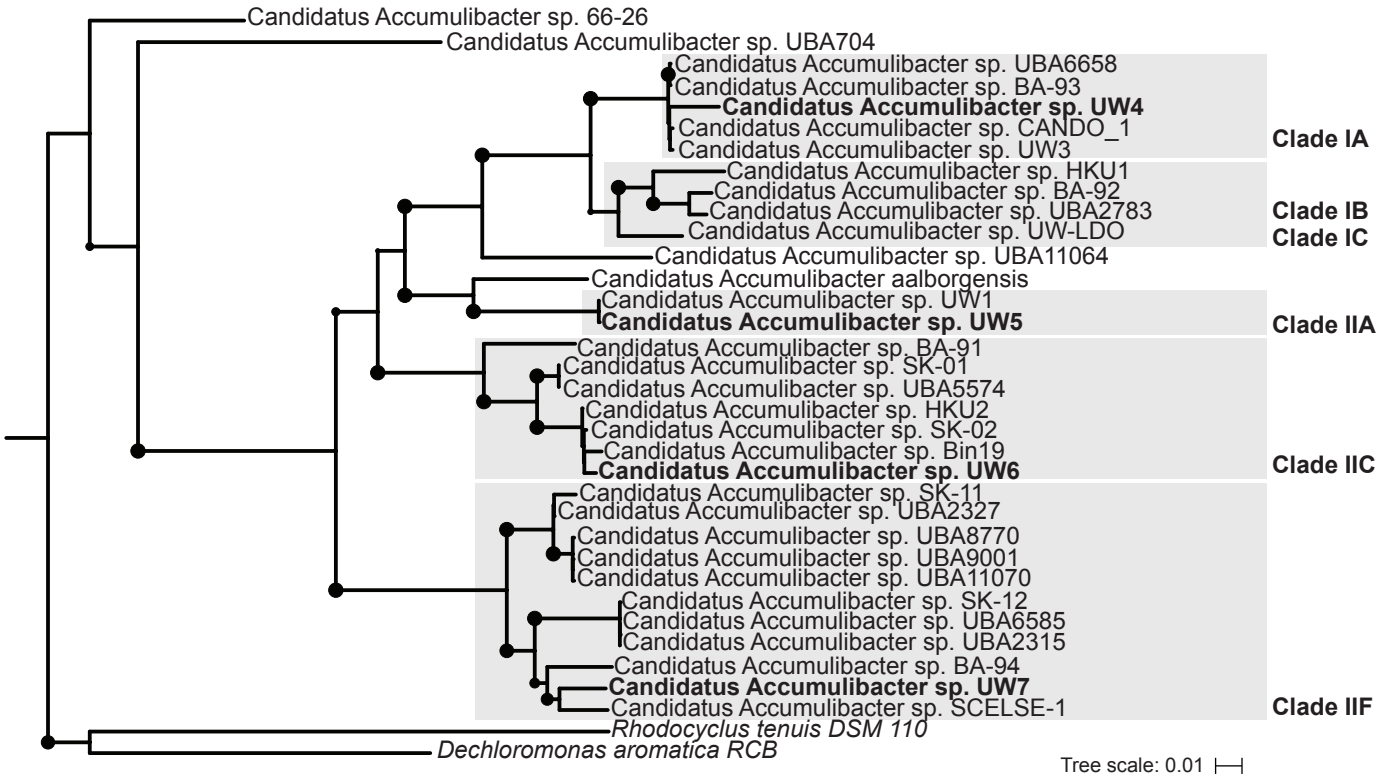

Anaerobic
Aerobic
2
1
0
-1
log2 fold-change
Presence of Orthologs:
Outgroup
Type I
Type II
Clade IA only
Protein CbbY
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein
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

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

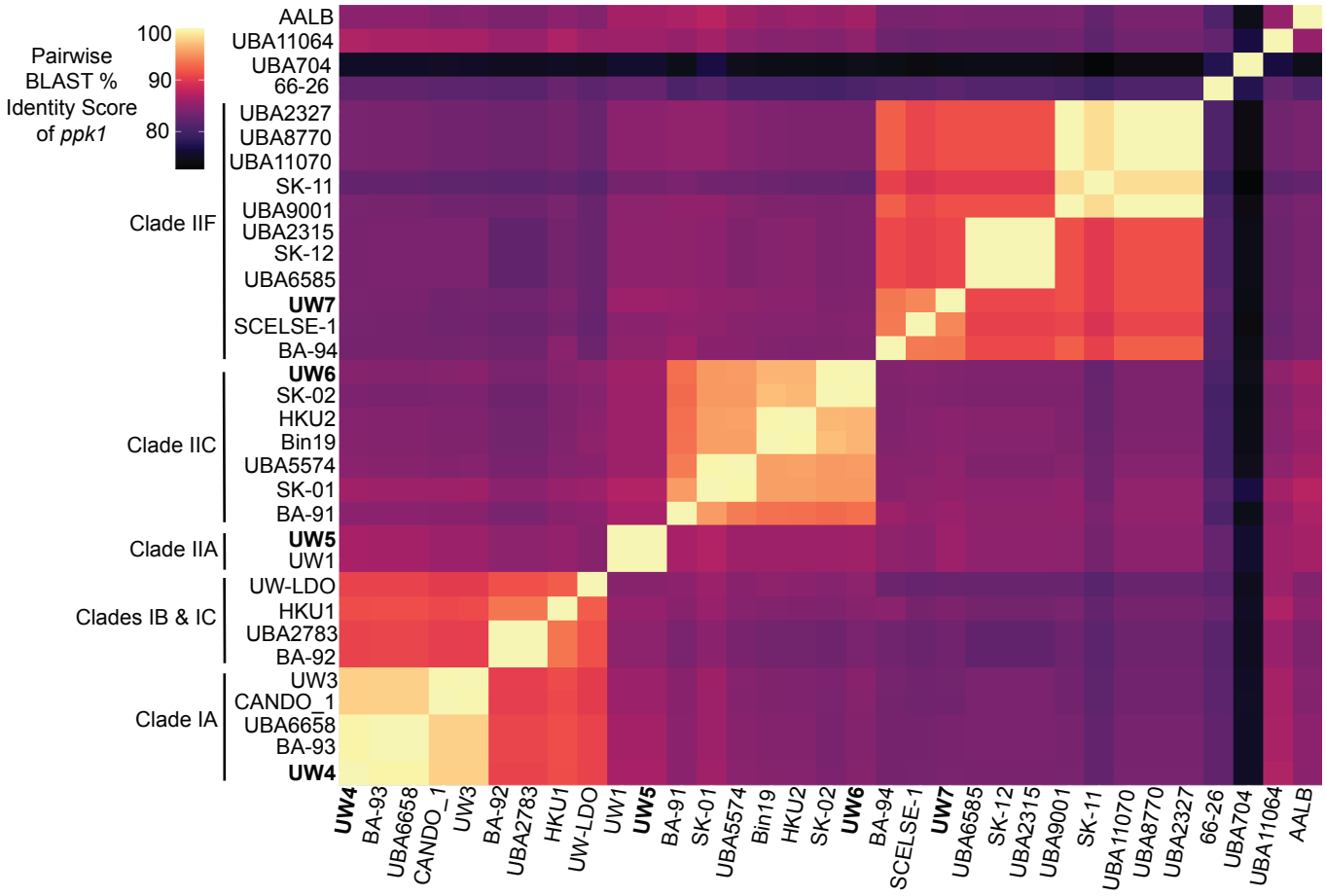
Clade Genome	Anaerobic-10:45	Anaerobic-11:16	Anaerobic-11:55	Aerobic-12:40	Aerobic-13:15	Aerobic-13:55	Aerobic-14:55	Total Reads Mapped
IIA-UW5	1063676	1357480	1706229	2197556	1763359	1171470	2065179	11324949
IIC-UW6	5226724	5892548	6755519	7902549	5004518	6914956	4132969	41829784
IIF-UW7	51038	78249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

	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