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], Pamela Y. Camejo[2,5], Benjamin O. Oyserman[6,7], Matthew Kizaric[2], Eng Hoe Khor[2], Daniel R. Noguera[2,8], 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] Genomics & Resistant Microbes (GeRM), Instituto de Ciencias e Innovación en Medicina, Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, 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] Bioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands

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

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

* equal contribution

† Corresponding Author: kdmcmahon@wisc.edu

ABSTRACT

Microbial communities in their natural habitats consist of closely related populations 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 clades 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 reads. We identified core and accessory genes to investigate how gene expression patterns differ among the dominating strains. Using this approach, we were able to identify putative pathways and functions that may differ between Accumulibacter clades 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 the
51 polyphosphate kinase (*ppk1)* locus sequence identity. These clades are predicted to have
52 key functional differences in acetate uptake rates, phage defense mechanisms, and
53 nitrogen cycling capabilities. However, such hypotheses have largely been made based
54 on gene-content comparisons of sequenced Accumulibacter genomes, some of which
55 were obtained from different systems. Here, we performed time-series genome-resolved
56 metatranscriptomics to explore gene expression patterns of co-existing Accumulibacter
57 clades in the same bioreactor ecosystem. Our work provides an approach for elucidating
58 ecologically relevant functions based on gene expression patterns between closely
59 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 Genetically diverse ecotypes of the marine cyanobacterium *Prochlorococcus* that are
75 97% identical by their 16S ribosomal RNA gene sequences exhibit markedly different
76 light-dependent physiologies and distinct seasonal and geographical patterns (1, 2, 4, 5).
77 Transcriptional-level differences between closely related microbial lineages have also
78 been demonstrated, such as that of high-light and low-light *Prochlorococcus* ecotypes
79 (6). However, dissecting the emergent ecological properties of genetic diversity observed
80 among coherent microbial lineages is challenging given that these principles are likely not
81 universal among the breadth of microbial diversity, and is further hindered by the lack of
82 pure cultures (7).

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 (8, 9).
89 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 are 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. Energy is recovered during aerobic PHA utilization by replenishing
97 polyphosphate and glycogen reserves, which can be later used for future PHA formation
98 as the anaerobic/aerobic cycle repeats. This feast-famine oscillation through sequential
99 anaerobic-aerobic cycles operating on minute-to-hourly time scales gives rise to net
100 phosphorus removal and the overall EBPR process (8, 10). A few 'omics-based studies
101 have revealed potentially regulatory modules that could coordinate Accumulibacter's
102 complex physiological behavior during these very dynamic environmental conditions (11–
103 14).

104 Accumulibacter can be subdivided into two main types (I and II) and further
105 subdivided into multiple clades based upon polyphosphate kinase (*ppk1*) sequence
106 identity, since the 16S rRNA marker is too highly conserved among the breadth of known
107 Accumulibacter lineages to resolve species-level differentiation (15–18). The two main
108 types share approximately 85% nucleotide identity across the *ppk1* locus (18) which
109 mirrors genome-wide average nucleotide identity (ANI) (19). Members of different clades
110 have been hypothesized to vary in their acetate uptake rates, nitrate reduction capacity,
111 and phage defense mechanisms (19, 20). For example, clade IA seems to exhibit higher
112 acetate uptake rates and phosphate release rates, whereas clade IIA can reduce nitrate
113 and clade IA cannot (21, 22). The persistence of several phylogenetically distinct clades
114 within the Accumulibacter lineage in both full-scale wastewater treatment plants and lab-

115 scale enrichment bioreactors suggests that different clades may inhabit distinct ecological
116 niches (23). This is largely supported by comparative genomics of numerous
117 Accumulibacter metagenome-assembled genomes (MAGs) recovered from different
118 bioreactor systems and classified using the *ppk1* locus (12, 19–21, 24). However, the
119 ecological roles that distinct Accumulibacter clades may play is not yet clear.

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

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 (24). 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. Allythiourea was added to the
151 synthetic feed solution to inhibit nitrification/denitrification.

152 Three biomass samples for metagenomic sequencing were collected in
153 September of 2015 by centrifuging 2-mL of mixed liquor at 8,000 *x g* for 2 min, and DNA
154 was extracted using a phenol:chloroform extraction protocol. Seven samples for RNA
155 sequencing were collected throughout a single EBPR cycle (3 in the anaerobic phase, 4
156 in the aerobic phase) in October of 2016 following the sampling strategy conducted by
157 Oyserman et al. (11). Samples were flash-frozen on liquid nitrogen immediately after
158 centrifuging and discarding the supernatant. RNA was extracted using a TRIzol-based
159 extraction method (Thermo Fisher Scientific, Waltham, MA) followed by phenol-
160 chloroform separation and RNA precipitation. RNA was purified following an on-column

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

166 **Metagenomic and Metatranscriptomic Library Construction and Sequencing**

167 Metagenomic libraries were prepared by shearing 100 ng of DNA to 550 bp long
168 products using the Covaris LE220 (Covaris) and size selected with SPRI beads (Beckman
169 Coulter). The fragments were ligated with end repair, A-tailing, Illumina compatible
170 adapters (IDT Inc) using the KAPA-Illumina library preparation kit (KAPA Biosystems).
171 Libraries were quantified using the KAPA Biosystem next-generation sequencing library
172 qPCR kit and ran on a Roche LightCycler 480 real-time PCR instrument. The quantified
173 libraries were prepared for sequencing on the Illumina HiSeq platform using the v4
174 TruSeq paired-end cluster kit and the Illumina cBot instrument to create a clustered flow
175 cell for sequencing. Shotgun sequencing was performed at the University of Wisconsin –
176 Madison Biotechnology Center DNA Sequencing facility on the Illumina HiSeq 2500
177 platform with the TruSeq SBS sequencing kits, followed by 2x150 indexing. Raw
178 metagenomic data consisted of 219.4 million 300-bp Illumina HiSeq reads with
179 approximately 3.9 Gpb per sample (Table 1 – supplementary material). Nanopore
180 sequencing was performed according to the user's manual version SQK-LSK108.

181 Total RNA submitted to the University of Wisconsin-Madison Biotechnology Center
182 was verified for purity and integrity using a NanoDrop 2000 Spectrophotometer and an
183 Agilent 2100 BioAnalyzer, respectively. RNA-Seq paired end libraries were prepared

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

197 **Metagenomic Assembly and Annotation**

198 We applied two different assembly and binning approaches to obtain high-quality
199 genomes of the dominant Accumulibacter strains present in the reactors. For the first
200 approach, Illumina unmerged reads were quality filtered and trimmed using the Sickle
201 software v1.33 (25). Reads were merged with FLASH v1.0.3 22 (26), with a mismatch
202 value of ≤0.25 and a minimum of ten overlapping bases from paired sequences, resulting
203 in merged read lengths of 150 to 290 bp. FASTQ files were then converted to FASTA
204 format using the Seqtk software v1.0 (27). Metagenomic reads from all three samples
205 were then co-assembled using the Velvet assembler with a k-mer size of 65 bp, a
206 minimum contig length of 200 bp, and a paired-end insert size of 300 bp (28). Metavelvet

207 was used to improve the assembly generated by Velvet (29). Metagenomic contigs were
208 then binned using Maxbin (30). This approach allowed us to assemble high-quality
209 genomes of clades IIA and IIC, named UW5 and UW6, respectively. Contigs from these
210 genomes were scaffolded using long Nanopore reads that were generated under the
211 standard protocol using MeDuSa (31), and manually inspected to remove short contigs
212 and decontaminated using ProDeGe (32) and Anvi'o (33). Further scaffolding was
213 performed on both of these genomes using Nanopore long reads using LINKS (34) and
214 GapCloser (35).

215 Because this pipeline did not produce high-quality genomes of clades IA and IIF,
216 we applied a different approach to assemble these genomes. Quality filtered reads from
217 each metagenome were individually assembled using metaSPAdes and co-assembled
218 together also using metaSPAdes (36). Reads from each metagenome were mapped
219 against each of the assemblies using bbmap with a 95% sequence identity cutoff to obtain
220 differential coverage (37), and contigs were binned using MetaBat (38). Identical clusters
221 of bins were dereplicated across assemblies using dRep (39) to obtain the best quality
222 genome of both clades IA and IIF, named UW4 and UW7, respectively. All genomes were
223 quality checked using CheckM (40) and functional annotations assigned with Prokka and
224 KofamKOALA (41, 42).

225 Each genome was assigned to a particular clade both by comparing the *ppk1*
226 sequence identity and ANI to previously published Accumulibacter genomes. A *ppk1*
227 database was created using sequences from previously published Accumulibacter
228 references and clone sequences. We searched for the *ppk1* gene in our draft genomes
229 using this database and aligned the corresponding hits with MAFFT (43). A phylogenetic

230 tree of aligned *ppk1* gene sequences from Accumulibacter references, select clone
231 sequences, and the outgroups *Dechloromonas aromatica RCB* and *Rhodc*yclus *tenuis*
232 *DSM 110* with RAxML v8.1 with 100 rapid bootstraps (44). To use any given
233 Accumulibacter reference genome in downstream comparisons, a confident *ppk1* hit had
234 to be identified using the above methods. As a technical note, the IA-UW2 strain
235 assembled by Flowers et al. (19) was renamed to UW3, as the original assembly
236 contained an additional contig that was likely from a prophage. The UW3 strain assembly
237 does not contain this contig, and therefore the UW numerical nomenclature follows in
238 sequential order after UW3. A phylogenetic tree of *ppk1* nucleotide sequences was
239 performed and compared to select clone sequences. Pairwise genome-wide ANI was
240 calculated between the four Accumulibacter draft genomes and all available
241 Accumulibacter genomes in Genbank with FastANI (45). A species tree of all
242 Accumulibacter references and outgroup genomes was constructed using single-copy
243 markers from the GTDBK-tk (46), aligned with MAFFT (43), built with RAxmL with 100
244 rapid bootstraps (44), and visualized in iTOL (47). To compare genome-wide ANI scores
245 of Accumulibacter references with pairwise nucleotide identity of the *ppk1* locus of all
246 references, we performed pairwise BLAST for all coding *ppk1* coding regions and
247 reported the percent identity (48).

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

249 We identified core and accessory gene content of Accumulibacter clades through
250 clustering orthologous groups of genes (COGs) using PyParanoid (49). We used the four
251 Accumulibacter MAGs generated in this study, high-quality Accumulibacter reference
252 genomes (above 90% completeness and less than 5% redundancy), and *Dechloromonas*

253 *aromatica RCB* and *Rhodocyclus tenuis DSM 110* as outgroups (Accumulibacter
254 references listed in Supplementary Table 1 available at
255 https://figshare.com/account/projects/90614/articles/13237148). Briefly, an all-vs-all
256 comparison of proteins for all genomes was performed using DIAMOND (50), pairwise
257 homology scores calculated with the InParanoid algorithm (51), gene families constructed
258 with MCL (52), and HMMs built for each gene family with HMMER (53). A phylogenetic
259 tree was constructed for genomes used for the COGs analysis by using the coding
260 regions of the PPK1 locus and overlaid with the presence/absence of core and flexible
261 gene content using the ete3 python package (54). A similarity score between UW
262 Accumulibacter genomes was calculated as the pairwise Pearson correlation coefficient
263 between gene families using the numpy python package (55). The presence and absence
264 of different gene families was visualized in an upset plot between UW Accumulibacter
265 genomes using the ComplexUpset package based on the UpSetR package in R (56). All
266 data wrangling and plotting was performed using the tidyverse suite of packages in R
267 (57).

268 **Metatranscriptomics Mapping and Processing**

269 Metatranscriptomics reads from each of the seven samples were quality filtered
270 using fastp (58) and rRNA removed with SortMeRNA (59). The coding regions of all four
271 Accumulibacter genomes assembled in this study were predicted with Prodigal (60) and
272 annotated with both Prokka and KofamKOALA (41, 42), and concatenated together to
273 create a mapping index. Metatranscriptomic reads were competitively pseudoaligned to
274 the reference index and counts quantified with kallisto (61). Genes with a sum of more
275 than 10,000 counts across all 7 samples were removed, as these are likely ribosomal

276 RNAs that were not removed previously. Additionally, genes annotated as rRNAs by
277 barrnap were manually removed (41). Low-count genes were removed by requiring that
278 all genes had to have more than 10 counts mapped across 3 or more samples.

279 To explore genes and metabolic functions that may exhibit hallmark anaerobic-
280 aerobic feast-famine cycling patterns, we identified sets of genes that were differentially
281 expressed between anaerobic and aerobic conditions using DESeq2 (62). Differential
282 expression calculations were performed for each strain separately to account for
283 differences in the relative abundance of metatranscriptomic reads mapping back to each
284 of the four MAGs. We analyzed expression patterns of differentially expressed genes
285 using a threshold cutoff of ± 1.5 log2 fold-change between anaerobic and aerobic
286 conditions and then identified differentially expressed genes as core or accessory among
287 clades IA (str. UW4) and IIC (str. UW6). For the core group analysis, the corresponding
288 ortholog in each genome had to meet the minimum count threshold as described above
289 to include in the comparison, with genes in one or both of the strains being differentially
290 expressed. For the accessory analysis for each strain, the gene had to be differentially
291 expressed and not contained in the other genome, but could be present in other strains
292 or other genomes belonging to that clade.

293 **Data and Code Availability**

294 Raw sequencing files for the three metagenomes and seven metatranscriptomes
295 have been submitted to NCBI under the BioProject accession PRJNA668760. Assembled
296 UW5 and UW6 genomes have been deposited through IMG at accession numbers
297 2767802316 and 2767802455 respectively. All genomes assemblies used in this study in
298 their current forms as well as transcriptomic count tables are available on Figshare at:

299 https://figshare.com/projects/Metabolic_Plasticity_of_Accumulibacter_Clades/90614. All
300 code is available at https://github.com/elizabethmcd/R3R4.
301
302 **RESULTS**
303 **Enrichment of Co-Existing Accumulibacter Clades**
304 We inoculated a 2-L sequencing-batch reactor with activated sludge from a full-
305 scale wastewater treatment plant in Madison, WI, USA and maintained it for 40 months.
306 The reactor was primarily fed with acetate and operated under cyclic anaerobic-aerobic
307 phases with a 4-day solids retention time to simulate the EBPR process and enrich for
308 Accumulibacter (Figure 1). During a period of stable EBPR (14 months following
309 inoculation) in which acetate was completely diminished by the start of the aerobic phase
310 and soluble phosphate was below 1.0 mg-L^{-1} at the end of the aerobic phase, we collected
311 seven samples across a single cycle (three in the anaerobic phase, four in the aerobic
312 phase) for RNA-sequencing (Figure 1A). We quantified the different Accumulibacter
313 clades by performing qPCR of the polyphosphate kinase (*ppk1*) locus as described
314 previously (23) from samples collected on the same day as the RNA-seq experiment
315 (Figure 1B). At the time of the RNA-seq experiment, the bioreactor was primarily enriched
316 in Accumulibacter clade IA, followed by clade IIC and clade IIA.
317 We assembled high-quality genomes of four Accumulibacter clades (Table 1),
318 including the dominant IA and IIC (UW4 and UW6, respectively) clades and less abundant
319 IIA and IIF clades (UW5 and UW7, respectively) (Figure 2). Each of the assembled
320 genomes falls within the established Accumulibacter clade nomenclature as defined by
321 *ppk1* sequence identity and phylogenetic placement when compared to other publicly

322 available Accumulibacter genomes and clone sequences (Figure 2A). Additionally, the
323 *ppk1* hierarchical structure is also reflected by pairwise-average nucleotide identity (ANI)
324 boundaries between clades, with some exceptions for newly assembled Accumulibacter
325 genomes that fall outside of the established '*C.* Accumulibacter phosphatis' lineage
326 (Figure 2B). All assembled genomes are above 90% complete and contain less than 5%
327 redundancy as calculated by CheckM (40) (Table 1). We were able to scaffold the
328 assemblies of strains IIA-UW5 and IIC-UW6 using long Nanopore reads, constructing a
329 completely contiguous draft genome of clade IIC. To our knowledge, the assembled clade
330 IIC genome (UW6) is the most contiguous, highest-quality reference genome available
331 for this clade, providing a valuable new Accumulibacter reference genome (Figure 2C).

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

333 We next characterized the functional diversity of different Accumulibacter clades
334 through clustering orthologous groups of genes (COGs) (Figure 3A). We performed an
335 updated COGs analysis from Skennerton et al. (20) and Oyserman et al. (63) to include
336 high-quality Accumulibacter references, and UW-generated genomes, and the outgroups
337 *Dechloromonas aromatica* and *Rhodocyclus tenuis* (Figure 3). The Accumulibacter
338 lineage exhibits a tremendous amount of functional diversity, with only approximately 25%
339 of COGs shared amongst all Accumulibacter clades (Figure 3A). Additionally, there is
340 extensive diversity within the two Accumulibacter types and within individual clades, with
341 only a few representatives of each clade more than approximately 75% similar to each
342 other by gene content (Supplementary Figure 1). Specifically, clade IIC broadly harbors
343 the most diversity of all sampled genomes by gene content, with representatives of this
344 clade not sharing more than 25% of COGs (Figure 3A). Other clades are more similar by

345 gene content, but this could be due to fewer genomes sampled from these clades. Most
346 of the highest-quality, publicly available Accumulibacter genome references belong to
347 clades IIF and IIC.

348 We then compared COGs between Accumulibacter clades collected from
349 bioreactors seeded from the Nine Springs Wastewater Treatment Plant in Madison, as
350 these genomes represent a shared origin (Figure 3B). In addition to two lab-scale
351 bioreactors harboring distinct Accumulibacter clades (genomes UW1 and UW3, and
352 UW4-7, respectively) (19, 24), we recently characterized clade IC (genome UW-LDO-IC)
353 using genome-resolved metagenomics and metatranscriptomics (12). By calculating the
354 Pearson correlation similarity of COG presence/absence, genomes in Type I and Type II
355 cluster respectively, where genomes within Type I are overall more similar to each other,
356 as expected (Supplementary Figure 1). Clade IIA genomes (strains UW1 and UW5) were
357 assembled from separate enrichments derived from the same full-scale treatment plant
358 approximately 12 years apart, but are very similar by gene families (Supplementary
359 Figure 1). UW genomes within Type II are less similar by gene content and seem to be
360 more diverse in this regard, though again this could be due to under-sampling of Type I
361 (8 genomes in Type I and 14 in Type II). Additionally, clade IIF contains more overlap with
362 Type I genomes in gene families than some Type II genomes do to each other
363 (Supplementary Figure 1).

364 COG overlap and uniqueness is shown by the intersection of groups of COGs
365 between UW Accumulibacter clades and the *Rhodocyclus* and *Decholoromonas*
366 outgroups (Figure 3B). There are 1075 gene families shared among all UW
367 Accumulibacter clades and outgroup genomes, with 277 separate gene families only

368 within the UW Accumulibacter clades (Figure 3B). Among individual clades, there are 887
369 groups only within clade IIC (UW6), 399 within clade IIA (UW1 and UW5), 399 within IIF
370 (UW7), 292 within clade IA (UW3 and UW4), and 196 within IC (UW-LDO). Although clade
371 IIC str. UW6 contains more overlap in gene content with SK-02 (Figure 3A), this
372 intersection analysis identifies gene families that are available to individual strains within
373 the same bioreactor ecosystem, highlighting the potential for functional differences. Below
374 we explore these putative functional differences in shared and flexible gene content
375 between Accumulibacter clades IA and IIC using time-series transcriptomics.

376 **Transcriptional Profiles of Accumulibacter Clades**

377 Although *ppk1* locus quantification showed that the bioreactor was mostly enriched
378 in clade IA (Figure 1B), surprisingly, more RNA-seq reads mapped to the clade IIC-UW6
379 genome than to the clade IA-UW4 (Figure 4A). Approximately 40 million total
380 transcriptional reads mapped back to the clade IIC-UW6 genome across the RNA-seq
381 experiment, whereas clades IA-UW4 and IIA-UW5 recruited roughly the same number of
382 reads, with 14 million and 11 million, respectively (Table 2). Markedly lower reads levels
383 mapped to the clade IIF-UW7 relative to the other three clades, and therefore we do not
384 include clade IIF in our subsequent analyses.

385 The co-existing clades in the bioreactor system differ in both the presence of sets
386 of nitrogen cycling gene sets and their expression profiles (Figure 4C). In clade IA-UW4,
387 we detected the *napAGH* genes for nitrate reduction, whereas we did not detect *napB*,
388 possibly due to these genes being close to the end of a contig in this genome. We also
389 detected a *nirS*-like nitrite reductase, both subunits of the nitric oxide reductase *norQD,*
390 and the nitrous oxide reductase *nosZ* in clade IA-UW4. Most of these genes in clade IA-

391 UW4 exhibit the highest expression patterns in the early and late anaerobic stages
392 (Figure 4C). Conversely, clade IIC-UW6 contains the *narGHIJ* machinery for nitrate
393 reduction, a different *nirS*-like nitrite reductase than in IA-UW4, and does not contain a
394 nitrous oxide reductase. However, clade IIC-UW6 does contain the nitrite reductase *nirBD*
395 for reducing nitrite to ammonium, which is uniformly expressed across the cycle (Figure
396 4C).

397 We explored the top differentially expressed genes in clade IIC-UW6, comparing
398 anaerobic and aerobic conditions (Figure 4B). Genes upregulated in the anaerobic phase
399 belonged to pathways for central-carbon transformations, fatty acid biosynthesis, and
400 PHA synthesis. Interestingly, all genes within the high-affinity phosphate transport system
401 Pst were upregulated in the aerobic phase. This includes the substrate-binding
402 component *pstS, phoU* regulator, and the ABC-type transporter *pstABC* complex. The
403 higher overall transcriptional levels of clade IIC-UW6 versus clade IA-UW4 (Figure 4A)
404 and upregulation of the high-affinity Pst system (Figure 4B) combined with the lower IIC-
405 UW6 abundance compared to IA-UW4, suggests that clade IIC-UW6 compensates for
406 lower abundance and competes for substrates through upregulation of transcriptional
407 programs related to the EBPR process.

408 **Expression Dynamics of Genes Shared by Clades IA and IIC**

409 We next characterized expression dynamics of COGs that are shared between
410 clades IA-UW4 and IIC-UW6 (Figure 5). There are very few orthologs that are
411 differentially expressed between the anaerobic and aerobic phases in both clades that
412 met the ±1.5-fold change threshold cutoff. These genes include a coproporphyrinogen III
413 oxidase, NAD-reducing HoxS subunits, the *phaC* subunit of the PHA synthase, the long-

414 chain fatty-acid-CoA ligase FadD13, and other hypothetical proteins. The fatty-acid CoA-
415 ligase incorporates ATP, CoA, and fatty acids of variable length to form acyl-CoA to be
416 degraded for energy production, incorporated into complex lipids, or used in other
417 metabolic pathways (64). The acyl-CoA ligase is upregulated in the anaerobic phase of
418 both clades, suggesting a role in anaerobic carbon metabolism, as acyl-CoA can
419 ultimately form acetyl-CoA through beta oxidation. The *phaC* subunit of the PHA synthase
420 is differentially expressed in the anaerobic phase for both clades, as is typical of the
421 hallmark EBPR metabolism for Accumulibacter (11, 65).

422 Additionally, both clades carry the bidirectional NiFe Hox system, encoding an
423 anaerobic hydrogenase. The HoxH and HoxY beta and delta subunits form the
424 hydrogenase moiety, and the FeS-containing HoxF and HoxU alpha and gamma subunits
425 catalyze NAD(P)H/NAD(P)+ oxidation/reduction coupled to the hydrogenase moiety (66–
426 68). In both clades, the FeS cluster HoxF and HoxU alpha and gamma subunits are
427 strongly differentially expressed in the anaerobic phases, as was observed previously in
428 a genome-resolved metatranscriptomic investigation of clade IIA-UW1 (11). However,
429 clade IIC-UW6 is missing the HoxY delta subunit of the hydrogenase moiety, and only
430 clade IA-UW4 exhibits differential expression of the HoxH beta subunit in the anaerobic
431 phase. Clade IIC-UW6 does exhibit higher expression of HoxH in the late anaerobic and
432 early aerobic phases, but does not meet the threshold cutoff for differential gene
433 expression. The prior study focused on clade IIA-UW1 also demonstrated hydrogenase
434 activity in the anaerobic phase, and also demonstrated hydrogen gas production after
435 acetate addition (11). The hydrogen gas production was hypothesized to replenish NAD+
436 after glycogen degradation. In this ecosystem, both clades may exhibit anaerobic

437 hydrogenase activity and thus produce hydrogen, but due to the missing subunit in clade
438 IIC-UW6 and lack of differential gene expression, this is uncertain.
439 Interestingly, several core ortholog subsets important for EBPR-related functions
440 exhibit differential expression patterns in only one of the clades. For example, subunits of
441 the high-affinity phosphate transporter Pst system are only differentially expressed in
442 clade IIC-UW6 and not IA-UW4, as observed for clade IIC-UW6 above (Figure 4C).
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 clade IA-UW4 and the denitrification protein NirQ is differentially expressed in
446 only clade IIC-UW6. Whereas the nitrogen fixation protein subunit NifU is differentially
447 expressed in both clades. 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. Although the clade IA-UW4 assembly is missing an acetate
450 kinase, a phosphate acetyltransferase is only differentially expressed in this clade. Within
451 clade IIC, both the acetate kinase and phosphate acetyltransferase genes exhibit
452 relatively low levels of expression over the time-series. Both clades contain several
453 acetyl-CoA synthases that are highly expressed in the anaerobic phase, particularly upon
454 acetate contact, but do not exhibit stark differential expression between the two phases.
455 Previous work based on community proteomics and enzymatic assays have
456 suggestedthat although both activation pathways are present and expressed in
457 Accumulibacter, the high-affinity pathway is preferentially used (69, 70). Although a
458 phosphate acetyltransferase is differentially expressed in clade IA, this gene has lower

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

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

462 We lastly characterized expression dynamics of flexible gene content between
463 clades IA-UW4 and IIC-UW6, focusing specifically on clade IA-UW4 (Figure 6). A majority
464 of the differentially expressed flexible genes in clade IIC-UW6 were annotated as
465 hypothetical proteins, and thus we focused on putative differentiating features of clade
466 IA-UW4. Flexible gene content for IA-UW4 can be defined as any ortholog that is not
467 present in the UW6 genome, whether that ortholog is only contained within clade IA
468 genomes, Type I as a whole, within other Type II or clade IIC genomes, or the
469 *Dechloromonas* or *Rhodocyclus* outgroups. Thus, the absence of a particular ortholog in
470 the clade IIC-UW6 genome but present in the IA-UW4 genome or other genomes was
471 inferred as a lack of functional capability of IIC in this particular ecosystem. For example,
472 this analysis identified several groups of genes that are annotated as hypothetical genes
473 that contain orthologs only within clade IA genomes and that are differentially expressed
474 between the anaerobic and aerobic cycles.

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

482 within the same COG, and thus clade IA-UW4 contains a different *nirS* that is upregulated
483 in the anaerobic phase. Clade IIC-UW6 does not contain both subunits of the nitric-oxide
484 reductase *norQD,* whereas clade IA-UW4 does. These subunits are not differentially
485 expressed above the set threshold but do exhibit the highest expression upon acetate
486 contact in the anaerobic phase (Figure 4B). Additionally, clade IIC does not contain the
487 nitrous-oxide reductase *nosZ* (Figure 4B), whereas clade IA does and this is differentially
488 expressed in the anaerobic phase (Figure 6). These results suggest that clade IA-UW4
489 may have the capability for full denitrification from nitrate to nitrogen gas, or at least use
490 the reduction of nitrate/nitrite as a source of energy in the anaerobic phase in a denitrifying
491 bioreactor. Although clade IIC-UW6 does contain the *nirBD* nitrite reductase complex for
492 reducing nitrite to ammonium, these genes are uniformly expressed throughout the cycle
493 (Figure 4B).
494
495 **DISCUSSION**
496 In this work, we performed time-series transcriptomics to explore gene expression
497 profiles of two Accumulibacter strains co-existing within the same bioreactor ecosystem.
498 By applying genome-resolved metagenomic and metatranscriptomic techniques, we were
499 able to postulate niche-differentiating features of distinct Accumulibacter clades based on
500 core and accessory gene content and differential expression profiles. We assembled
501 high-quality genomes from Accumulibacter clades, using long reads to generate a single
502 contiguous scaffold of the IIC-UW6 genome. By performing an updated clustering of
503 orthologous groups of genes, we were able to better demonstrate the similarity and
504 uniqueness among Accumulibacter strains from this system. Unexpectedly, although

505 clade IA dominates in abundance by *ppk1* locus quantification, clade IIC is
506 transcriptionally more active. By comparing the expression profiles of COGs shared
507 between clades IA-UW4 and IIC-UW6, we found that genes involved in EBPR-related
508 pathways may differ between the clades with respect to transcription-level regulation.
509 Notably, clade IIC-UW6 shows a strong upregulation of the high-affinity Pit phosphorus
510 transport system in the aerobic phase and both clades made hydrogenase transcripts in
511 the anaerobic phase. Clade IA-UW4 contains most of the subunits for the individual key
512 steps for denitrification, and exhibits strong anaerobic upregulation of these genes,
513 suggesting a differentiating role for clade IA-UW4 in nitrogen cycling.

514 The Accumulibacter lineage harbors extensive diversity between and amongst
515 individual clades which is exhibited not only by *ppk1*-based phylogenies, but also by
516 pairwise genome-wide ANI and COG comparisons. For example, clade IIA-UW1 and
517 clade IA-UW3 were present in the same bioreactor ecosystem and are only similar by
518 85% ANI (19). In the bioreactor enrichment from this study, the dominant clades IA-UW4
519 and IIC-UW6 are less than 80% similar by genome-wide ANI (Figure 2B). Recently, a
520 cutoff of >95% ANI has been suggested to delineate distinct bacterial species or cohesive
521 genetic units (45), originally based on observations of coverage gaps at 90-95%
522 sequence identity when metagenomic reads were mapped back to assembled MAGs
523 (71–74). These sequence cutoffs have also been benchmarked against homologous
524 recombination rates and ratios of nonsynonymous to synonymous (*dN/dS*) nucleotide
525 differences (75). However, there are some notable exceptions to the 95% ANI threshold
526 cutoff used for delineating species, as distinct populations of the marine SAR11 clade
527 have been observed at a 92% boundary and lower (76, 77). Among Accumulibacter

528 genomes adhering to the established *ppk1*-based phylogeny, genomes within individual
529 clades fall within the >95% ANI threshold, whereas ANI comparisons within and between
530 Types I and II are more variable (Figure 2B). Genomes within Type I are similar by
531 approximately 90% ANI, whereas genomes within Type II are not as coherent.

532 The apparent coexistence of diverse Accumulibacter clades within an ecosystem
533 then raises questions of ecological roles, putative interactions, and evolutionary dynamics
534 of this lineage over time and space. Although the concept of a bacterial species is highly
535 contested and dependent upon the operational definition of the population in question,
536 species have been clustered based on phenotypic and metabolic characteristics, and
537 more recently by DNA-based clustering metrics and recombination signatures (3, 7). The
538 ecotype model of speciation suggests that cohesive, ecologically distinct ecotypes can
539 coexist because they occupy separate niche spaces (3, 78). Because Accumulibacter
540 has not been isolated in pure culture to date, most ecological and evolutionary inferences
541 have been made from genome-based comparisons (20, 63) or mapping metagenomic
542 reads back to assembled genomes (79, 80). However, the apparent diversity within the
543 Accumulibacter lineage may require reevaluation of the phylogeny and taxonomy overall.
544 Although the *ppk1*-based clade designations coincide with genome-wide ANI based
545 cutoffs remarkably well (Supplementary Figure 2B), there are some discrepancies. A few
546 Accumulibacter genomes have been assembled that do not adhere to the *ppk1*-based
547 population structure, but fall within the Accumulibacter lineage based on a species tree
548 of single-copy core genes (Supplementary Figure 2A), such as the proposed novel
549 species '*Candidatus* Accumulibacter aalborgensis' (81). Additionally, a few genomes
550 preliminarily classified by the GTDB as Accumulibacter branch outside of the established

551 nomenclature, and are quite divergent based on the species and *ppk1* phylogenies as
552 well as genome-wide and *ppk1* sequence similarities (Figure 2B, Supplementary Figure
553 2). Overall, substantial work is needed to reevaluate the phylogenetic structure of the
554 Accumulibacter lineage as a whole and connect with signatures of homologous
555 recombination and evolutionary trajectories within coexisting clades to understand the
556 genetic boundaries and thus ecological interdependencies of individual clades or strains.
557 Although the boundaries of individual Accumulibacter clades or strains have not
558 been entirely resolved to infer separate ecotypes, denitrification capabilities of specific
559 clades has been repeatedly suggested to be a niche-differentiating feature (12, 21, 23).
560 Previously, two separate bioreactors fed with either acetate or propionate as the sole
561 carbon source were enriched with different Accumulibacter morphotypes, one with cocci
562 morphology and the other rod morphology, which were also linked to different
563 denitrification capabilities (22), A bioreactor highly enriched in clade IC inferred
564 simultaneous use of oxygen, nitrite, and nitrate as electron acceptors under micro-aerobic
565 conditions either due to metabolic flexibility or heterogeneity within the IC population (23).
566 The resident strain IC-UW-LDO also possessed the full suite of genes for denitrification
567 and expressed them in microaerobic conditions (12); whereas other clades either do not
568 contain the complete genetic repertoire for denitrification or have not been shown to
569 express them. As we observed differential expression of denitrification genes in clade IA-
570 UW4 but not clade IIC-UW6 in this study, these expression differences could explain why
571 the two clades can coexist and contribute to overall ecosystem functioning (e.g.
572 phosphorus removal) by using different electron acceptors in the anaerobic/aerobic
573 conditions. Although the synthetic feed for the bioreactors in this study inhibits

574 nitrification/denitrification through alythiourea addition, the expression profiles of
575 denitrification genes in clade IA-UW4 suggests this strain could have a differentiating role
576 in a denitrifying bioreactor. Recently, a denitrifying EBPR bioreactor enriched in clade IA
577 encoded genes for complete denitrification, whereas the co-occurring clade IC strain and
578 other flanking community members did not (82). Additionally, Gao et al. performed an
579 ancestral genome reconstruction analysis similar to Oyserman et al. 2017 (63) to
580 specifically understand the flux of denitrification gene families. Interestingly, the *nir* gene
581 cluster was identified as a core COG that is core to all sampled Accumulibacter genomes,
582 whereas the periplasmic nitrate reductase *napAGH* and nitrous oxide reductase *nosZDFL*
583 clusters were core to only Type I Accumulibacter genomes, and inferred to have been
584 lost for some Type II genomes (82). We observed differential expression of nitrate
585 reductase and nitrous-oxide reductase genes in clade IA-UW4, where homologs of these
586 genes were not present in the clade IIC-UW6 genome and other Type II genomes.

587 Accumulibacter's anaerobic metabolism has also been shown to differ under
588 varying substrate or stoichiometric conditions and possibly between the two
589 Accumulibacter types. From recent simulations of 'omics datasets paired with enzymatic
590 assays, the routes of anaerobic metabolism for Accumulibacter may differ based on
591 environmental conditions (83). Under high acetate conditions, polyphosphate
592 accumulation arises through the glyoxylate shunt, whereas under increased glycogen
593 concentrations, glycolysis is used in conjunction with the reductive branch of the TCA
594 cycle (83). Furthermore, Welles et al. found that Type II Accumulibacter is able to switch
595 to partial glycogen degradation more quickly than Type I, enabling Type II to fuel VFA
596 uptake in polyphosphate-limited conditions (84, 85). Conversely, when concentrations of

597 polyphosphate are high, VFA uptake rates are higher in Type I. However, these studies
598 only identified Accumulibacter lineages based on ppk1 types, and therefore extrapolated
599 metabolic assumptions from presumably more resolved strains to the entire type. These
600 broad interpretations may obscure apparent functional differences within the two types
601 and individual clades. Future experiments could integrate multi-'omics approaches to
602 investigate not only metabolic flexibility under different substrate conditions and
603 perturbations, but how metabolisms are distributed amongst Accumulibacter clades in
604 bioreactor ecosystems that are not completely enriched in a single clade or strain.
605

606 **CONCLUSIONS**

607 In this study, we integrated genome-resolved metagenomics and time-series
608 metatranscriptomics to understand gene expression patterns of co-existing
609 Accumulibacter clades within a bioreactor ecosystem. We found evidence for
610 denitrification gene expression in the dominating clade IA, but higher overall
611 transcriptional activity for clade IIC along with differential expression of a high-affinity
612 phosphorus transporter. These results highlight the need for reevaluating the
613 phylogenetic structure of the Accumulibacter lineage, and defining how potential genetic
614 boundaries may translate to individual ecological niches allowing for the coexistence of
615 multiple clades or strains. Additionally, we highlight an approach for exploring functions
616 of closely related, uncultivated microbial lineages by exploring gene expression patterns
617 of shared and accessory genes. Overall, this work underscores the need to understand
618 the basic ecology and evolution of microorganisms underpinning biotechnological

processes to better control treatment process and resource recovery outcomes in the future (86).

ACKNOWLEDGEMENTS

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

642 **REFERENCES**

643 1. Moore LR, Rocap G, Chisholm SW. 1998. Physiology and molecular phytogeny of
644 coexisting Prochlorococcus ecotypes. Nature 393:464–467.
645 2. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding
646 H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R,
647 Chisholm SW. 2014. Single-cell genomics reveals hundreds of coexisting
648 subpopulations in wild Prochlorococcus. Science (80-) 344:416–420.
649 3. Cohan FM. 2001. Bacterial Species and Speciation. Syst Biol 50:513–524.
650 4. N K, SE R, JW B-T, M G, R S, SW C. 2018. Fundamental differences in diversity
651 and genomic population structure between Atlantic and Pacific Prochlorococcus.
652 .
653 5. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW.
654 2006. Niche partitioning among Prochlorococcus ecotypes along ocean-scale
655 environmental gradients. Science (80-) 311:1737–1740.
656 6. Voigt K, Sharma CM, Mitschke J, Joke Lambrecht S, Voß B, Hess WR, Steglich
657 C. 2014. Comparative transcriptomics of two environmentally relevant
658 cyanobacteria reveals unexpected transcriptome diversity. ISME J 8:2056–2068.
659 7. Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species
660 challenge: Making sense of genetic and ecological diversity. Science (80-).
661 Science.
662 8. Hesselmann RPX, Werlen C, Hahn D, van der Meer JR, Zehnder AJB. 1999.
663 Enrichment, Phylogenetic Analysis and Detection of a Bacterium That Performs
664 Enhanced Biological Phosphate Removal in Activated Sludge. Syst Appl Microbiol

665 22:454–465.
666 9. He S, Mcmahon KD. 2011. Microbiology of “Candidatus Accumulibacter” in
667 activated sludge. Microb Biotechnol 4:603–619.
668 10. Seviour RJ, Mino T, Onuki M. 2003. The microbiology of biological phosphorus
669 removal in activated sludge systems. FEMS Microbiol Rev 27:99–127.
670 11. Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. 2016.
671 Metatranscriptomic insights on gene expression and regulatory controls in
672 Candidatus Accumulibacter phosphatis. ISME J 10:810–822.
673 12. Camejo PY, Oyserman BO, McMahon KD, Noguera DR. 2019. Integrated Omic
674 Analyses Provide Evidence that a “Candidatus Accumulibacter phosphatis” Strain
675 Performs Denitrification under Microaerobic Conditions. mSystems 4:e00193-18.
676 13. He S, Kunin V, Haynes M, Martin HG, Ivanova N, Rohwer F, Hugenholtz P,
677 McMahon KD. 2010. Metatranscriptomic array analysis of “Candidatus
678 Accumulibacter phosphatis”-enriched enhanced biological phosphorus removal
679 sludge. Environ Microbiol 12:1205–1217.
680 14. Law Y, Kirkegaard RH, Cokro AA, Liu X, Arumugam K, Xie C, Stokholm-
681 Bjerregaard M, Drautz-Moses DI, Nielsen PH, Wuertz S, Williams RBH. 2016.
682 Integrative microbial community analysis reveals full-scale enhanced biological
683 phosphorus removal under tropical conditions. Sci Rep 6:1–15.
684 15. Mcmahon KD, Dojka M a, Pace NR, Jenkins D, Keasling JD. 2002.
685 Polyphosphate Kinase from Activated Sludge Performing Enhanced Biological
686 Phosphorus Removal Polyphosphate Kinase from Activated Sludge Performing
687 Enhanced Biological Phosphorus Removal †. Appl Environ Microbiol 68:4971–

688 4978.
689 16. He S, Gall DL, McMahon KD. 2007. “Candidatus accumulibacter” population
690 structure in enhanced biological phosphorus removal sludges as revealed by
691 polyphosphate kinase genes. Appl Environ Microbiol 73:5865–5874.
692 17. Peterson SB, Warnecke F, Madejska J, McMahon KD, Hugenholtz P. 2008.
693 Environmental distribution and population biology of Candidatus Accumulibacter,
694 a primary agent of biological phosphorus removal. Environ Microbiol 10:2692–
695 2703.
696 18. McMahon KD, Yilmaz S, He S, Gall DL, Jenkins D, Keasling JD. 2007.
697 Polyphosphate kinase genes from full-scale activated sludge plants. Appl
698 Microbiol Biotechnol 77:167–173.
699 19. Flowers JJ, He S, Malfatti S, del Rio TG, Tringe SG, Hugenholtz P, McMahon KD.
700 2013. Comparative genomics of two “Candidatus Accumulibacter” clades
701 performing biological phosphorus removal. ISME J 7:2301–2314.
702 20. Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. 2015. Expanding our
703 view of genomic diversity in *C andidatus* Accumulibacter clades. Environ
704 Microbiol 17:1574–1585.
705 21. Flowers JJ, He S, Yilmaz S, Noguera DR, McMahon KD. 2009. Denitrification
706 capabilities of two biological phosphorus removal sludges dominated by different
707 “Candidatus Accumulibacter” clades. Environ Microbiol Rep 1:583–588.
708 22. Carvalho G, Lemos PC, Oehmen A, Reis MAM. 2007. Denitrifying phosphorus
709 removal: Linking the process performance with the microbial community structure.
710 Water Res 41:4383–4396.

711 23. Camejo PY, Owen BR, Martirano J, Ma J, Kapoor V, Santo Domingo J, McMahon
712 KD, Noguera DR. 2016. Candidatus Accumulibacter phosphatis clades enriched
713 under cyclic anaerobic and microaerobic conditions simultaneously use different
714 electron acceptors. Water Res 102:125–137.
715 24. Martín HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C,
716 He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL,
717 Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P. 2006.
718 Metagenomic analysis of two enhanced biological phosphorus removal (EBPR)
719 sludge communities. Nat Biotechnol 24:1263–1269.
720 25. Joshi N, Fass J. 2011. Sickle: A sliding-window, adaptive, quality-based trimming
721 tool for FastQ files.
722 26. Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to
723 improve genome assemblies. Bioinformatics 27:2957–2963.
724 27. Li H. 2013. Seqtk: a fast and lightweight tool for processing FASTA or FASTQ
725 sequences.
726 28. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly
727 using de Bruijn graphs. Genome Res 18:821–829.
728 29. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: An extension of
729 Velvet assembler to de novo metagenome assembly from short sequence reads.
730 Nucleic Acids Res 40.
731 30. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an
732 automated binning method to recover individual genomes from metagenomes
733 using an expectation-maximization algorithm. Microbiome 2:26.

734 31. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R,
735 Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–
736 51.
737 32. Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl
738 JL, Ivanova N, Woyke T, Kyrpides N, Pati A. 2016. ProDeGe: A computational
739 protocol for fully automated decontamination of genomes. ISME J 10:269–272.
740 33. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.
741 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data.
742 PeerJ 3:e1319.
743 34. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I.
744 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long
745 reads. Gigascience 4:35.
746 35. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T, Sun
747 N, Rao J, Wang J, Zhang W, Chen Y, Liao S, Jiang H, Liu X, Yang Z, Mu F, Gao
748 S. 2017. LR_Gapcloser: a tiling path-based gap closer that uses long reads to
749 complete genome assembly. Gigascience 6:1.
750 36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
751 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V, Sirotkin A V, Vyahhi N, Tesler
752 G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly
753 algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–77.
754 37. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read
755 merging via overlap. PLoS One 12:e0185056.
756 38. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for

757 accurately reconstructing single genomes from complex microbial communities.
758 PeerJ 3:e1165.
759 39. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: A tool for fast and
760 accurate genomic comparisons that enables improved genome recovery from
761 metagenomes through de-replication. ISME J 11:2864–2868.
762 40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
763 assessing the quality of microbial genomes recovered from isolates, single cells,
764 and metagenomes. Genome Res 25:1043–55.
765 41. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
766 30:2068–2069.
767 42. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H.
768 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and
769 adaptive score threshold. bioRxiv 602110.
770 43. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid
771 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
772 30:3059–66.
773 44. Stamatakis A. 2014. The RAxML v8 . 0 . X Manual 1–55.
774 45. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
775 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
776 boundaries. Nat Commun 9:5114.
777 46. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to
778 classify genomes with the Genome Taxonomy Database. Bioinformatics.
779 47. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the

780 display and annotation of phylogenetic and other trees. Nucleic Acids Res
781 44:W242–W245.
782 48. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden
783 TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
784 49. Melnyk RA, Hossain SS, Haney CH. 2019. Convergent gain and loss of genomic
785 islands drive lifestyle changes in plant-associated Pseudomonas. ISME J
786 13:1575–1588.
787 50. Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using
788 DIAMOND. Nat Methods. Nature Publishing Group.
789 51. Sonnhammer ELL, Östlund G. 2015. InParanoid 8: Orthology analysis between
790 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43:D234–D239.
791 52. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-
792 scale detection of protein families. Nucleic Acids Res 30:1575–84.
793 53. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol
794 7:e1002195.
795 54. Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and
796 Visualization of Phylogenomic Data. Mol Biol Evol 33:1635–1638.
797 55. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D,
798 Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson
799 J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ,
800 Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R,
801 Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F,
802 van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner

803 A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C,
804 Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V., Olivetti E, Martin E,
805 Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen
806 GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J,
807 Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV,
808 Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K,
809 Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ,
810 Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT,
811 Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T,
812 Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T,
813 Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. 2020. SciPy 1.0:
814 fundamental algorithms for scientific computing in Python. Nat Methods 17:261–
815 272.
816 56. Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the
817 visualization of intersecting sets and their properties. Bioinformatics 33:2938–
818 2940.
819 57. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund
820 G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K,
821 Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo
822 K, Yutani H. 2019. Welcome to the Tidyverse. J Open Source Softw 4:1686.
823 58. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
824 preprocessor. Bioinformatics 34:i884–i890.
825 59. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of

826 ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.
827 60. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010.
828 Prodigal: prokaryotic gene recognition and translation initiation site identification.
829 BMC Bioinformatics 11:119.
830 61. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-
831 seq quantification. Nat Biotechnol 34:525–527.
832 62. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and
833 dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.
834 63. Oyserman BO, Moya F, Lawson CE, Garcia AL, Vogt M, Heffernen M, Noguera
835 DR, McMahon KD. 2016. Ancestral genome reconstruction identifies the
836 evolutionary basis for trait acquisition in polyphosphate accumulating bacteria.
837 ISME J 10:2931–2945.
838 64. Watkins PA. 2018. The Fatty Acyl-CoA SynthetasesReference Module in
839 Biomedical Sciences. Elsevier.
840 65. He S, McMahon KD. 2011. “Candidatus Accumulibacter” gene expression in
841 response to dynamic EBPR conditions. ISME J 5:329–340.
842 66. Long M, Liu J, Chen Z, Bleijlevens B, Roseboom W, Albracht SPJ. 2007.
843 Characterization of a HoxEFUYH type of [NiFe] hydrogenase from
844 Allochromatium vinosum and some EPR and IR properties of the hydrogenase
845 module. J Biol Inorg Chem 12:62–78.
846 67. Antal TK, Oliveira P, Lindblad P. 2006. The bidirectional hydrogenase in the
847 cyanobacterium Synechocystis sp. strain PCC 6803. Int J Hydrogen Energy
848 31:1439–1444.

849 68. Vignais PM, Billoud B. 2007. Occurrence, classification, and biological function of
850 hydrogenases: An overview. Chem Rev. Chem Rev.
851 69. Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, Hettich RL,
852 Bond PL, VerBerkmoes NC, Banfield JF. 2008. Community proteogenomics
853 highlights microbial strain-variant protein expression within activated sludge
854 performing enhanced biological phosphorus removal. ISME J 2:853–864.
855 70. Hesselmann RPX, Von Rummell R, Resnick SM, Hany R, Zehnder AJB. 2000.
856 Anaerobic metabolism of bacteria performing enhanced biological phosphate
857 removal. Water Res 34:3487–3494.
858 71. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J,
859 Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG,
860 Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR.
861 2016. Genome-wide selective sweeps and gene-specific sweeps in natural
862 bacterial populations. ISME J 10:1589–1601.
863 72. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T,
864 Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018.
865 Contrasting patterns of genome-level diversity across distinct co-occurring
866 bacterial populations. ISME J 12:742–755.
867 73. Caro-Quintero A, Konstantinidis KT. 2012. Bacterial species may exist,
868 metagenomics reveal. Environ Microbiol 14:347–355.
869 74. Konstantinidis KT, DeLong EF. 2008. Genomic patterns of recombination, clonal
870 divergence and environment in marine microbial populations. ISME J 2:1052–
871 1065.

872 75. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield
873 JF. 2020. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial
874 Species Boundaries. mSystems 5.
875 76. Tsementzi D, Wu J, Deutsch S, Nath S, Rodriguez-R LM, Burns AS, Ranjan P,
876 Sarode N, Malmstrom RR, Padilla CC, Stone BK, Bristow LA, Larsen M, Glass
877 JB, Thamdrup B, Woyke T, Konstantinidis KT, Stewart FJ. 2016. SAR11 bacteria
878 linked to ocean anoxia and nitrogen loss. Nature 536:179–183.
879 77. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, Giovannoni S, Eren
880 AM. 2019. Single-amino acid variants reveal evolutionary processes that shape
881 the biogeography of a global SAR11 subclade. Elife 8.
882 78. Cohan FM. 2006. Towards a conceptual and operational union of bacterial
883 systematics, ecology, and evolution, p. 1985–1996. *In* Philosophical Transactions
884 of the Royal Society B: Biological Sciences. Royal Society.
885 79. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, Cordero
886 OX. 2018. Strain-level diversity drives alternative community types in millimetre-
887 scale granular biofilms. Nat Microbiol 3:1295–1303.
888 80. Albertsen M, Benedicte L, Hansen S, Saunders AM, Nielsen H, Lehmann Nielsen
889 K. 2011. A metagenome of a full-scale microbial community carrying out
890 enhanced biological phosphorus removal. ISME J 6:1094–1106.
891 81. Albertsen M, McIlroy SJ, Stokholm-Bjerregaard M, Karst SM, Nielsen PH. 2016.
892 “Candidatus Propionivibrio aalborgensis”: A novel glycogen accumulating
893 organism abundant in full-scale enhanced biological phosphorus removal plants.
894 Front Microbiol 7:1033.

895 82. Gao H, Mao Y, Zhao X, Liu WT, Zhang T, Wells G. 2019. Genome-centric
896 metagenomics resolves microbial diversity and prevalent truncated denitrification
897 pathways in a denitrifying PAO-enriched bioprocess. Water Res 155:275–287.
898 83. Guedes da Silva L, Olavarria Gamez K, Castro Gomes J, Akkermans K, Welles L,
899 Abbas B, van Loosdrecht MCM, Wahl SA. 2020. Revealing the metabolic
900 flexibility of Candidatus Accumulibacter phosphatis through redox cofactor
901 analysis and metabolic network modeling. Appl Environ Microbiol.
902 84. Welles L, Abbas B, Sorokin DY, Lopez-Vazquez CM, Hooijmans CM, van
903 Loosdrecht MCM, Brdjanovic D. 2017. Metabolic Response of “Candidatus
904 Accumulibacter Phosphatis” Clade II C to Changes in Influent P/C Ratio. Front
905 Microbiol 7:2121.
906 85. Welles L, Tian WD, Saad S, Abbas B, Lopez-Vazquez CM, Hooijmans CM, van
907 Loosdrecht MCM, Brdjanovic D. 2015. Accumulibacter clades Type I and II
908 performing kinetically different glycogen-accumulating organisms metabolisms for
909 anaerobic substrate uptake. Water Res 83:354–366.
910 86. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE,
911 O’Malley MA, García Martín H, Pfleger BF, Raskin L, Venturelli OS, Weissbrodt
912 DG, Noguera DR, McMahon KD. 2019. Common principles and best practices for
913 engineering microbiomes. Nat Rev Microbiol. Nature Publishing Group.
914
915
916
917

918 **FIGURE AND TABLE LEGENDS**

919

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

921 **1A.** Overview of EBPR cycle and sampling scheme. Samples were collected for RNA
922 sequencing over the course of a normal EBPR cycle, with reported measurements for
923 soluble phosphorus and acetate normalized by volatile suspended solids (VSS). Samples
924 were taken at time-points 0, 31, 70, 115, 150, 190, and 250 minutes. **1B.** Abundance of
925 different Accumulibacter clades by *ppk1* copies per ng of DNA. Quantification of
926 Accumulibacter clades was performed by qPCR of the *ppk1* gene as described in Camejo
927 et al. (23).

928

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

930 **2A.** Phylogenetic tree of *ppk1* nucleotide sequences from existing Accumulibacter
931 references, clone sequences, and the four assembled Accumulibacter genomes from this
932 study. Tips in bold represent *ppk1* sequences from metagenome-assembled genomes,
933 whereas others are sequences from clone fragments available at the referenced
934 accession numbers. Starred tips represent *ppk1* sequences from the four assembled
935 Accumulibacter genomes from this study. Tree was constructed using RaXML with 100
936 rapid bootstraps. **2B.** Pairwise genome-wide ANI comparisons of all publicly available
937 Accumulibacter references and the four new Accumulibacter genomes, denoted in bold.
938 Genomes are grouped according to clade designation by *ppk1* sequence identity and in
939 order of the *ppk1* phylogeny, except for those that do not fall in an established clade. **2C.**
940 Genome diagram of the UW6 Accumulibacter IIC MAG. Layers represent from top to

941 bottom: coverage over 1000 bp windows, GC content and skew across the same 1000
942 bp windows, coding sequences on the positive and negative strands, and positions of
943 predicted rRNA (red) and tRNA (black) sequences.
944
945 **Figure 3. Core and Flexible Gene Content among the Accumulibacter lineage and**
946 **UW genomes**
947 **3A.** Phylogenetic tree of the coding regions for the PPK1 locus among high-quality
948 Accumulibacter references, UW genomes, and outgroup taxa *Dechloromonas aromatica*
949 *RCB* and *Rhodocyclus tenuis DSM 110.* Pie chart at each node represents the fraction of
950 genes that are shared and flexible for genomes belonging to that clade. **3B.** Upset plot
951 representing the intersections of groups of orthologous gene clusters among UW
952 genomes and outgroups, analogous to a Venn diagram. Each bar represents the number
953 of orthologous gene clusters, and the dot plot represents the genomes in which the groups
954 intersect.
955
956 **Figure 4. Transcriptomic Profiles of Accumulibacter Clades**
957 **4A.** Average expression profiles of each reference genome in the anaerobic and aerobic
958 phases. Reads were competitively mapped to the four assembled Accumulibacter clade
959 genomes and normalized by transcripts per million (TPM). The sum of the total counts of
960 the three samples in the anaerobic phase and four samples in the aerobic phase were
961 averaged, respectively, and plotted on a $\log_{10}$ scale. **4B.** Top 50 differentially-expressed
962 genes in clade IIC-UW6. All 50 genes pass the threshold of $\pm$1.5 log-10 fold-change
963 across the cycle. Annotations were predicted from a combination of Prokka (41) and

964 KofamKOALA (42). **4C.** Denitrification gene expression among clades IIC-UW6, IA-UW4,
965 and IIA-UW5. We summarized expression for genes involved in denitrification by
966 highlighting parts of the cycle in which certain genes are expressed the most. This is
967 characterized by acetate contact (pink), early anaerobic (dark purple), late anaerobic
968 (purple), early aerobic (dark blue) and late aerobic (light blue), where uniform describes
969 uniform expression across the entire cycle (grey). If an arrow is missing for a step, that
970 particular clade did not contain confident annotations within the genome for those genes.
971 For nitrate reduction, the *nar* system is represented as a black dot and the *nap* system is
972 represented as a grey dot. Multiple arrows are shown for a single step in which particular
973 genes were highly expressed in multiple phases.
974
975 **Figure 5. Differential Expression of Shared Genes in Clades IA and IIC**
976 Gene expression profiles of core COGs across the EBPR cycle that are differentially
977 expressed between anaerobic and aerobic conditions in either IA-UW4 or IIC-UW6, or
978 both strains. To consider a gene differentially expressed in a strain, the COG must exhibit
979 greater or less than ±1.5 fold-change in expression. Homologs that are directly compared
980 to each other in IA-UW4 and IIC-UW6 are orthologous genes that belong to the same
981 cluster.
982
983 **Figure 6. Differential Expression of Accessory Genes in Clade IA**
984 Gene expression profiles of groups of accessory gene groups in clade IA-UW4 that are
985 not present in the UW6-IIC genome. A gene was considered differentially expressed
986 between the anaerobic and aerobic cycles if it exhibited greater or less than ±1.5 fold

987 change in expression. Colored dots represent whether that gene contains orthologs within
988 outgroup genomes, other Type II Accumulibacter other than the UW6-IIC genome, within
989 Type I Accumulibacter, or only within clade IA.
990
991 **Table 1. Assembled Accumulibacter Genome Statistics**
992 Genome quality calculations for completeness and contaminations were made with
993 CheckM based on the presence/absence of single copy genes (40).
994
995 **Table 2. Metatranscriptomic Reads Mapped to Accumullibacter Genomes**
996 Counts of raw metatranscriptomic reads mapped to each Accumulibacter genome with
997 kallisto (61) in each sample for the three anaerobic samples and four aerobic samples,
998 and total reads mapped for all samples.
999
1000 **Supplementary Figure 1.**
1001 Pairwise Pearson correlation coefficient of shared gene content as a measure of how
1002 similar each of the UW genomes are to each other based on COGs.
1003
1004 **Supplementary Figure 2.**
1005 **S2A)** Species tree of all Accumulibacter reference genomes, UW-assembled genomes,
1006 and outgroup genomes using single-copy markers from the GTDB-tk (46). Phylogenetic
1007 tree was constructed using RAxmL with 100 rapid bootstraps and visualized in iTOL. **S2B)**
1008 Pairwise BLAST percent identity scores for *ppk1* coding regions for each reference
1009 genome in Figure 2B, in order of the *ppk1* phylogeny in Figure 2A. Raw pairwise percent

1010 identity scores are available at

1011 https://figshare.com/articles/dataset/Pairwise_Accumulibacter_ppk1_BLAST_Percent_D

1012 _Scores/13237247.

1013

1014 **Supplementary Table 1**

1015 Genome accessions and statistics for all Accumulibacter references used in phylogenies,

1016 ANI comparisons, and COG calculations. A reference genome was used for COG

1017 comparisons if it was a high quality genome (>95% complete and <5% redundant).

1018 Available at https://figshare.com/account/projects/90614/articles/13237148

1019

1020 **Supplementary File 1**

1021 Raw pairwise genome-wide ANI results for each genome calculated with FastANI.

1022 Available at https://figshare.com/account/projects/90614/articles/13237244

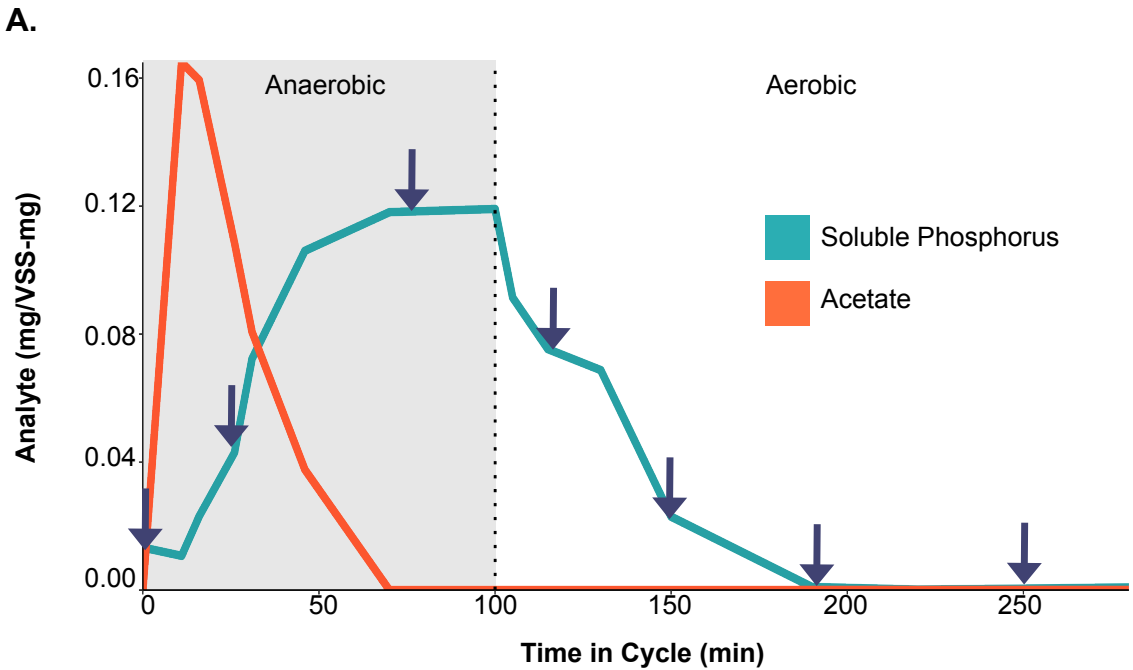
1023

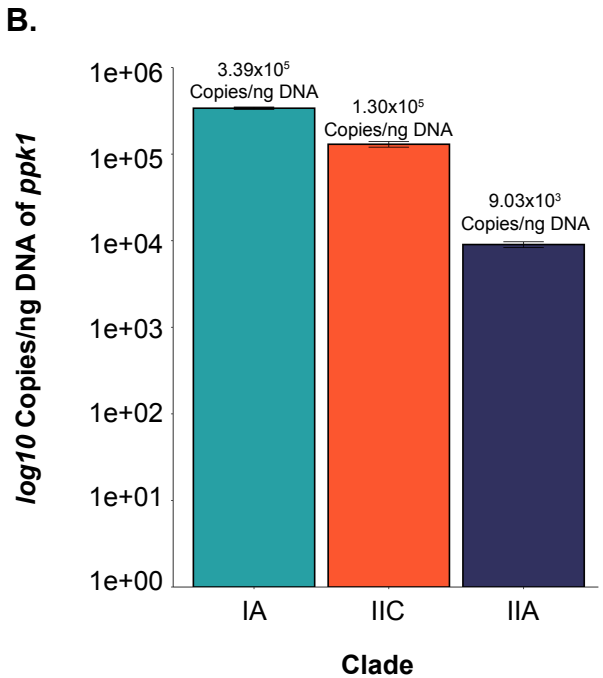
1024 **Supplementary File 2**

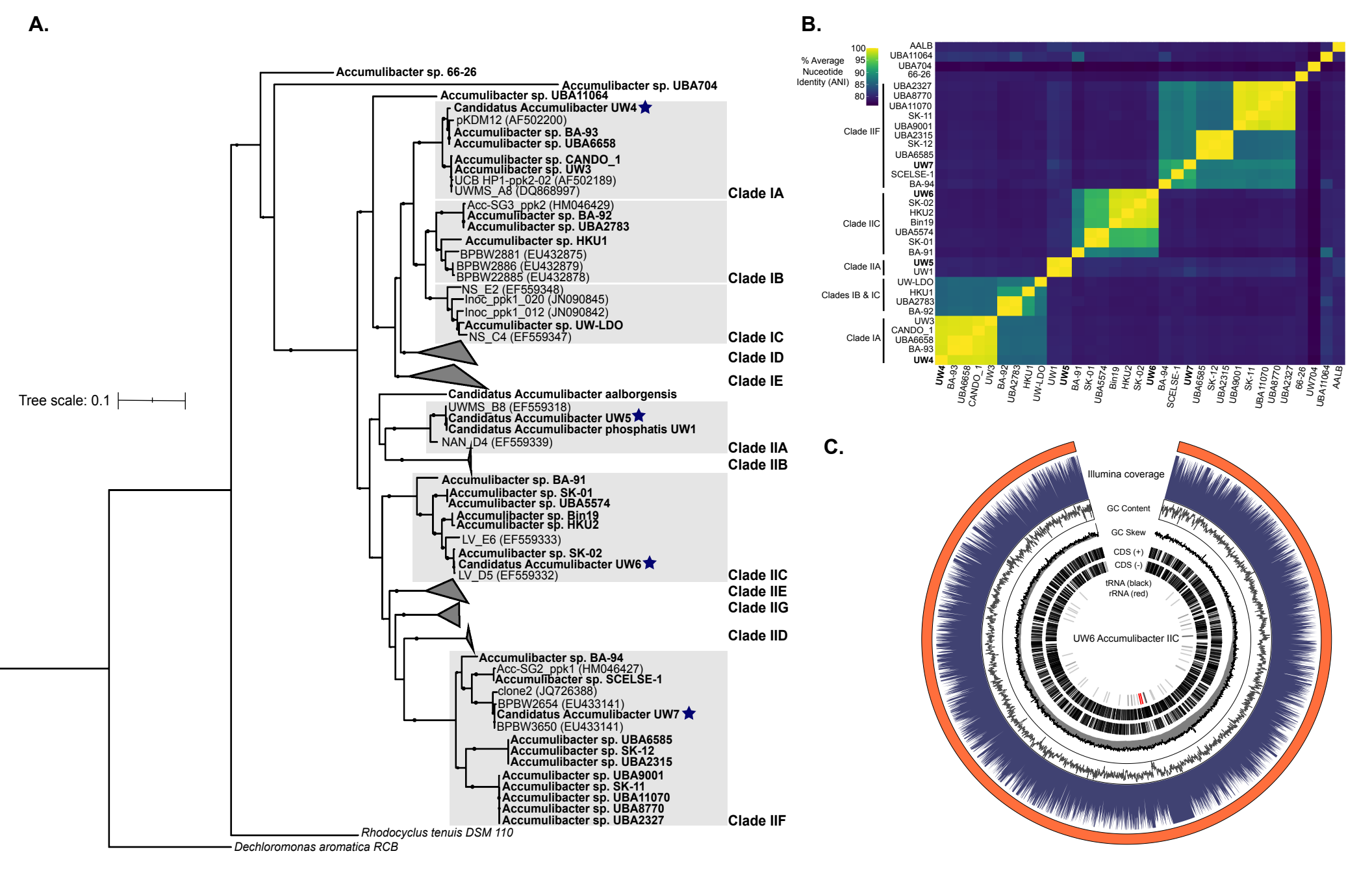
1025 Raw pairwise BLAST percent identity (PID) results for *ppk1* genes from each reference

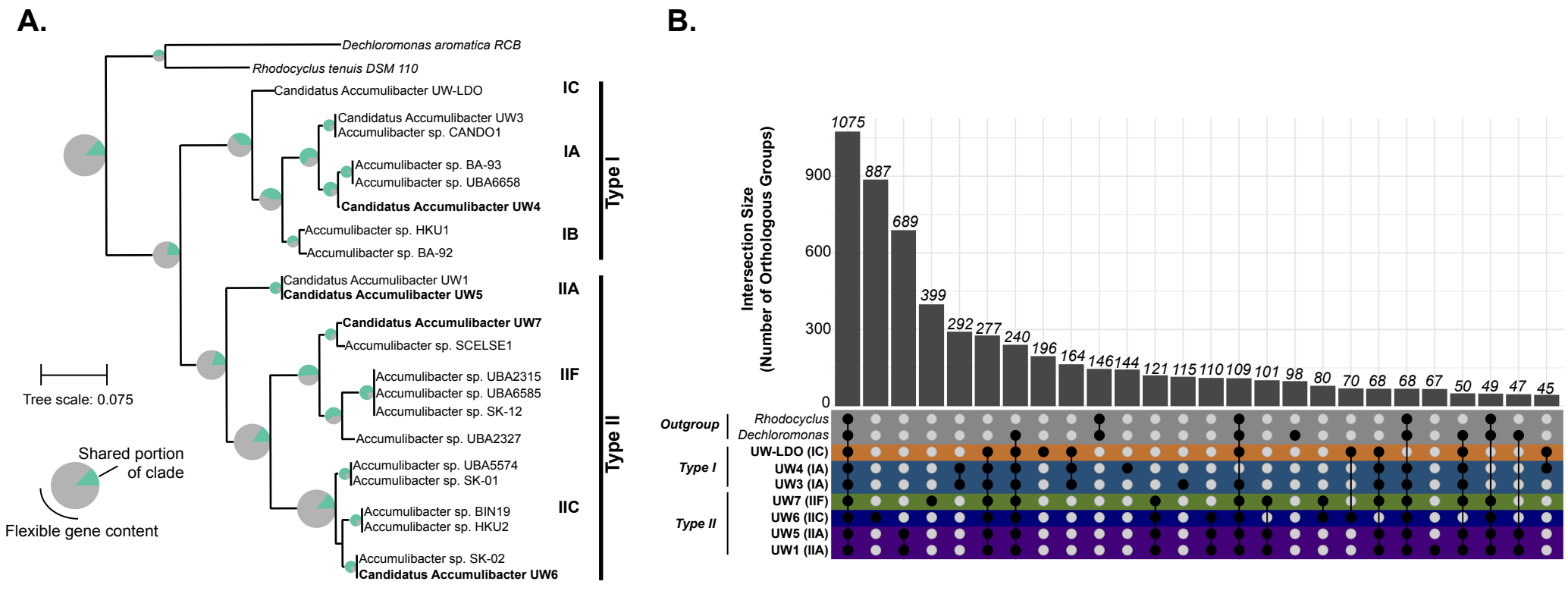
1026 genome. Available at https://figshare.com/account/projects/90614/articles/13237247

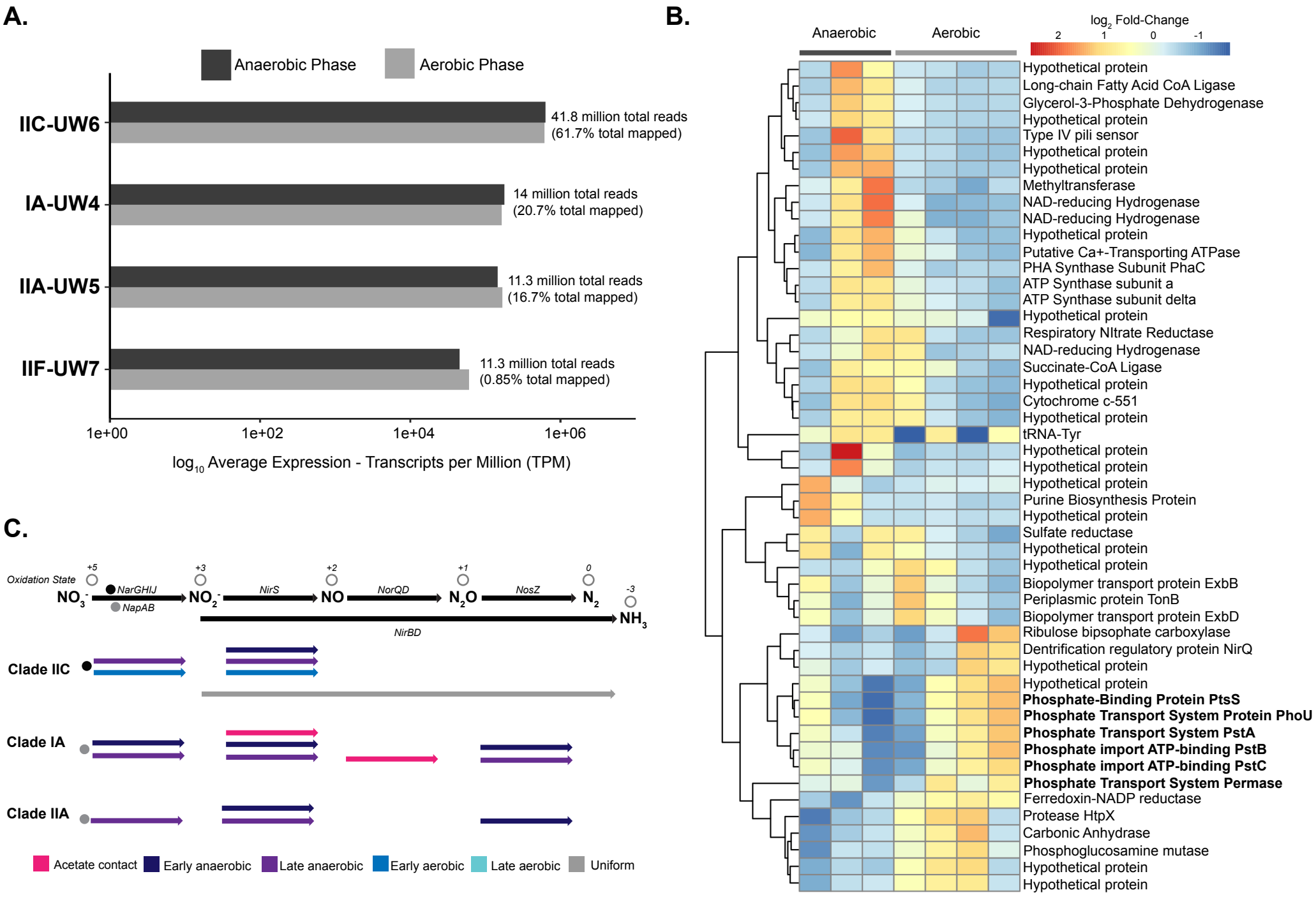
1027

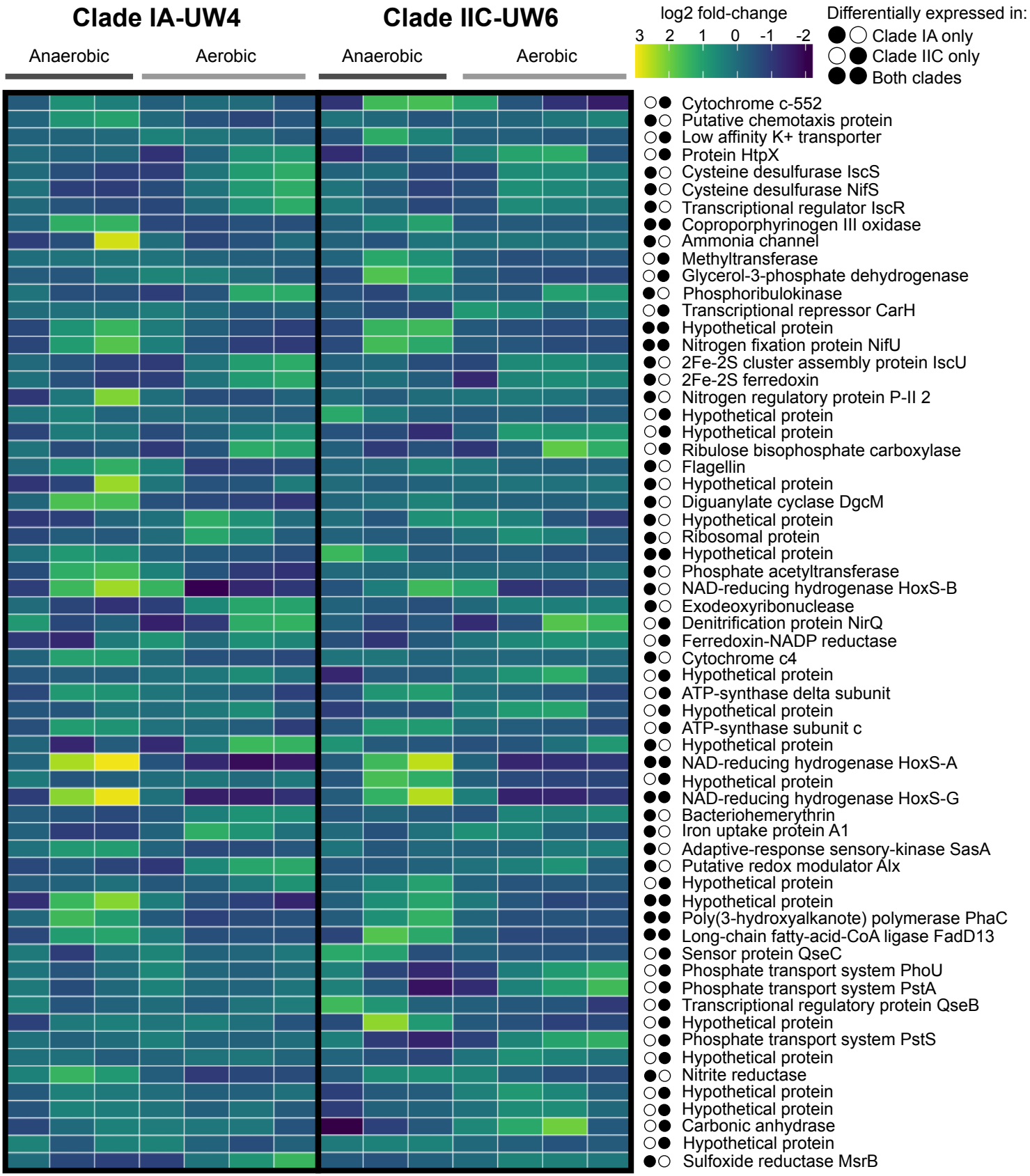

A.
Anaerobic
Aerobic
Soluble Phosphorus
Acetate
Analyte (mg/VSS-mg)
0.16
0.12
0.08
0.04
0.00
0
50
100
150
200
250
Time in Cycle (min)

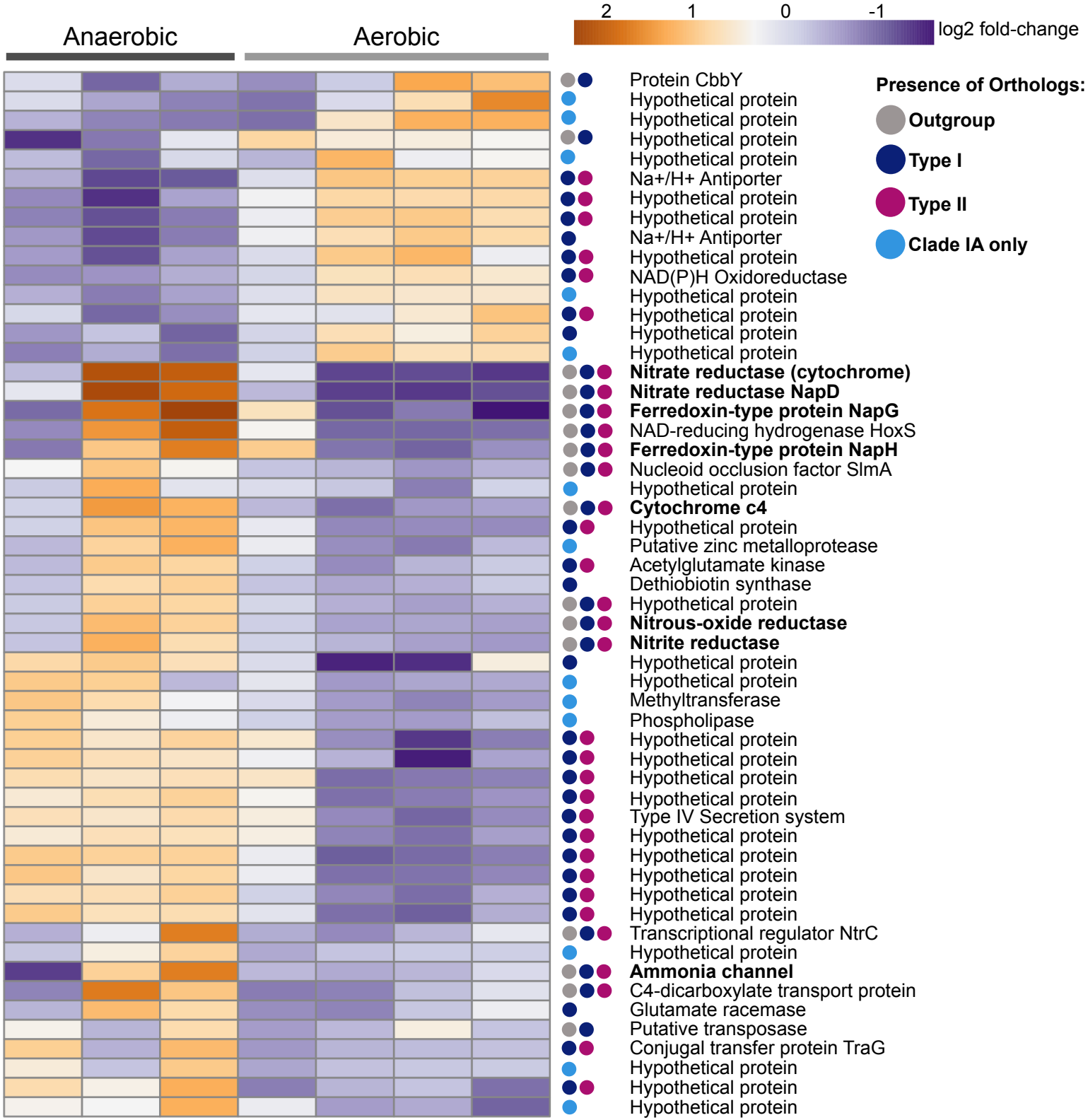

B.
3.39x10^5
Copies/ng DNA
1.30x10^5
Copies/ng DNA
9.03x10^3
Copies/ng DNA
log10 Copies/ng DNA of ppk1
1e+06
1e+05
1e+04
1e+03
1e+02
1e+01
1e+00
IA
IIC
IIA
Clade


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
11.3 million total reads
(0.85% total mapped)
1e+00
1e+02
1e+04
1e+06
log10 Average Expression - Transcripts per Million (TPM)
B.
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 PtsS
Phosphate Transport System Protein PhoU
Phosphate Transport System PstA
Phosphate import ATP-binding PstB
Phosphate import ATP-binding PstC
Phosphate Transport System Permase
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
NH3
NirBD
Clade IIC
Clade IA
Clade IIA
Acetate contact
Early anaerobic
Late anaerobic
Early aerobic
Late aerobic
Uniform


Clade IA-UW4
Clade 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

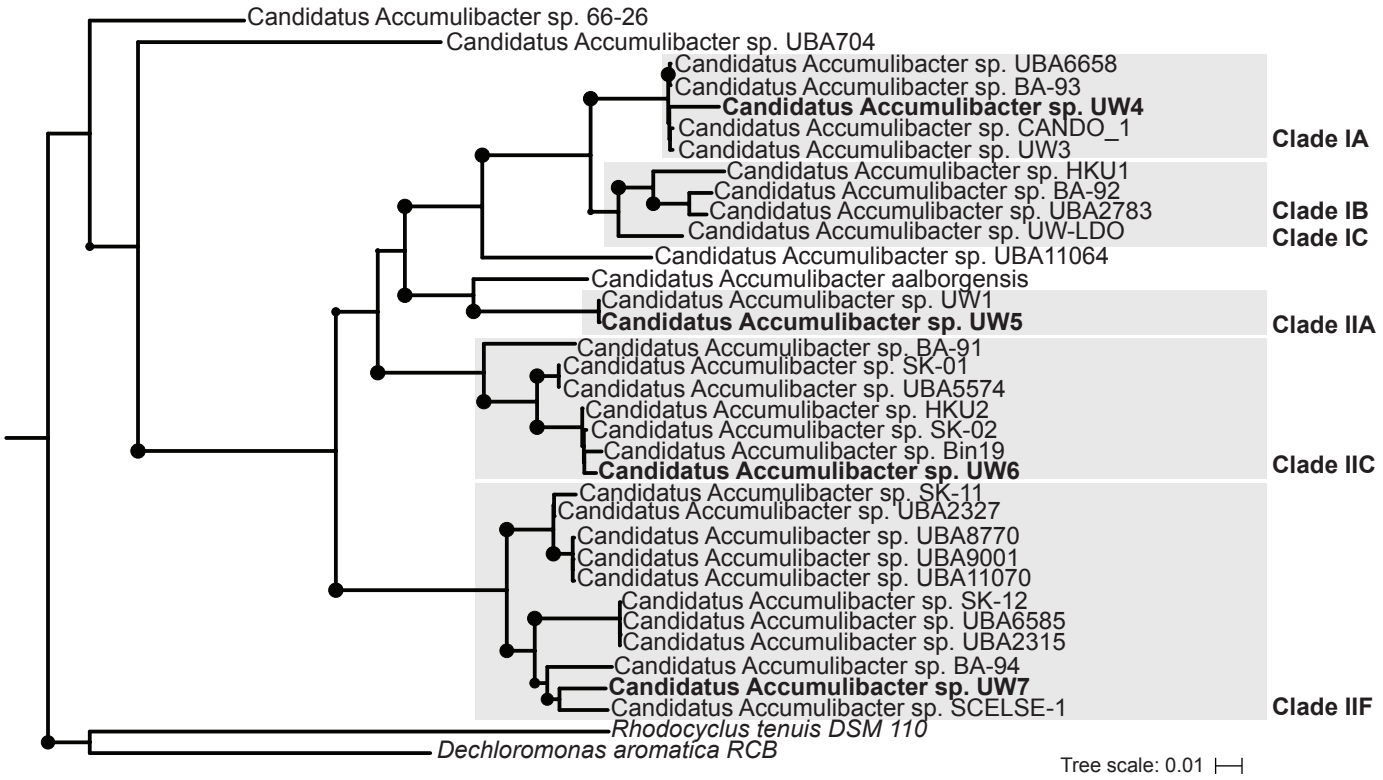

Anaerobic
Aerobic
2
1
0
-1
log2 fold-change
Presence of Orthologs:
Outgroup
Type I
Type II
Clade IA only
Protein CbbY
Hypothetical protein
Hypothetical protein
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein
Hypothetical protein
Na+/H+ Antiporter
Hypothetical protein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

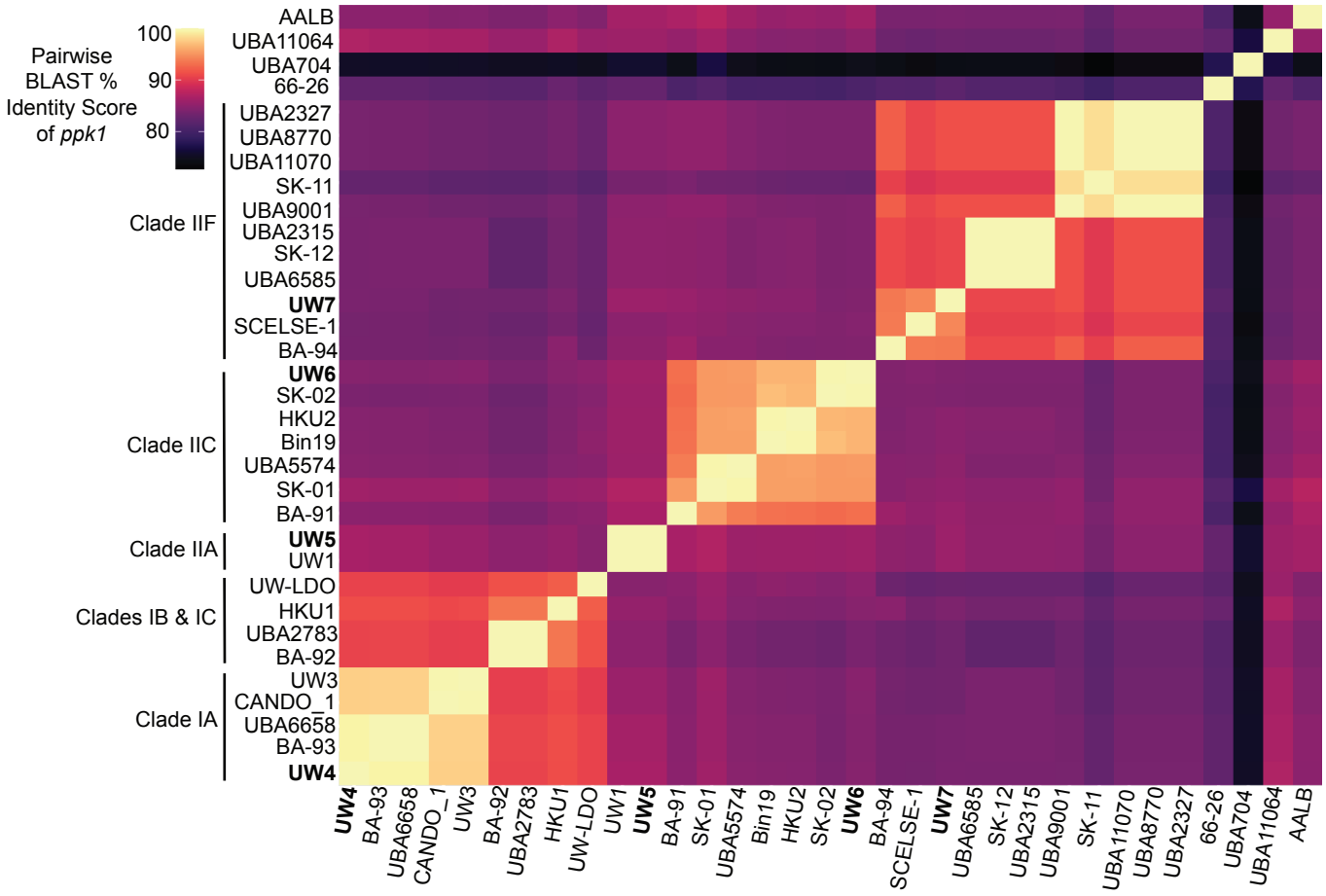
Clade Genome	Anaerobic-10:45	Anaerobic-11:16	Anaerobic-11:55	Aerobic-12:40	Aerobic-13:15	Aerobic-13:55	Aerobic-14:55	Total Reads Mapped
IIA-UW5	1063676	1357480	1706229	2197556	1763359	1171470	2065179	11324949
IIC-UW6	5226724	5892548	6755519	7902549	5004518	6914956	4132969	41829784
IIF-UW7	51038	78249	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