1   **Oligotyping and Genome-Resolved Metagenomics Reveal Distinct *Candidatus***

2   **Accumulibacter Communities in Full-Scale Side-Stream versus Conventional Enhanced**

3   **Biological Phosphorus Removal (EBPR) Configurations**

4   **Running Title:** Oligotyping reveals clade differences of Accumulibacter

5   Varun N. Srinivasan[1], Guangyu Li[1], Dongqi Wang[1,2], Nicholas B. Tooker[1,3], Zihan Dai[4],
6   Annalisa Onnis-Hayden[1], Ameet Pinto[1], April Z. Gu[1,5*]

7   [1]Department of Civil and Environmental Engineering, Northeastern University, Boston MA-
8   02115, USA

9   [2]State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of
10   Technology, Xi'an, Shaanxi 710048, China

11
12   [3]Department of Civil and Environmental Engineering, University of Massachusetts-Amherst,
13   Amherst MA-01002, USA

14
15   [4]Infrastructure and Environment Division, University of Glasgow, Glasgow G12 8LT, United
16   Kingdom

17   [5]Civil and Environmental Engineering, Cornell University, Ithaca NY – 14853, USA

18   [*]**Corresponding Authors**

19   April Gu, Civil and Environmental Engineering, Cornell University (aprilgu@cornell.edu)

20

Abstract

Candidatus Accumulibacter phosphatis (CAP) and its sub-clades-level diversity has been associated and implicated in successful phosphorus removal performance in enhanced biological phosphorus removal (EBPR). Development of high-throughput untargeted methods to characterize clades of CAP in EBPR communities can enable a better understanding of Accumulibacter ecology at a higher-resolution beyond OTU-level in wastewater resource recovery facilities (WRRFs). In this study, for the first time, using integrated 16S rRNA gene sequencing, oligotyping and genome-resolved metagenomics, we were able to reveal clade-level differences in Accumulibacter communities and associate the differences with two different full-scale EBPR configurations. The results led to the identification and characterization of a distinct and dominant Accumulibacter oligotype - Oligotype 2 (belonging to Clade IIC) and its matching MAG (RC14) associated with side-stream EBPR configuration. We are also able to extract MAGs belonging to CAP clades IIB (RCAB4-2) and II (RC18) which did not have representative genomes before. This study demonstrates and validates the use of a high-throughput approach of oligotyping analysis of 16S rRNA gene sequences to elucidate CAP clade-level diversity. We also show the existence of a previously uncharacterized diversity of CAP clades in full-scale EBPR communities through extraction of MAGs, for the first time from full-scale facilities.

Introduction

Enhanced biological phosphorus removal (EBPR) has been a promising technology for phosphorus removal from municipal wastewater, due to its sustainable nature and P recovery potential in comparison to chemical precipitation and adsorption-based approaches [1]. However, conventional EBPR processes face challenges related to system stability and susceptibility to

44   variations and fluctuations in influent loading conditions such as unfavorable carbon to P (C/P)

45   ratios [2]. A new emerging process, namely, side-stream EBPR (S2EBPR) features an anaerobic

46   side-stream reactor that allows a portion of the return activated sludge (RAS) to undergo

47   hydrolysis and fermentation thus enabling influent carbon-independent enrichment and selection

48   of polyphosphate accumulation organisms (PAO). Facilities that have operated or piloted

49   S2EBPR showed improved and more stable performance [3–6]. However, several key

50   knowledge gaps still exist for this process including lack of understanding of the fundamental

51   mechanisms governing the process and differences in microbial ecology, particularly

52   functionally relevant key populations such as PAOs between the conventional and the S2EBPR

53   processes. Several studies that have characterized the microbial ecology of different process

54   configurations using 16S rRNA gene amplicon sequencing, fluorescence in-situ hybridization

55   (FISH) and quantitative polymerase chain reaction (qPCR) have failed to associate differences in

56   PAO microbial ecology with process configurations [7–10]. This is, at least partially due to the

57   challenges in identification and association of specific PAO populations with various full-scale

58   EBPR configurations. The complex and overwhelming "masking" effect of facility-specific

59   selection forces exerted on the microbial community structure makes it difficult to isolate the

60   effect of process configuration alone.

61   *Candidatus* (*Ca.*) Accumulibacter phosphatis (hereafter referred to as Accumulibacter or CAP) is

62   considered to be a key PAO in both lab-scale and full-scale EBPR systems [11–16]. Our current

63   understanding of the physiology, metabolic potential, transcriptional, and proteomic dynamics

64   are based on lab-scale Accumulibacter-enriched reactors from which 19 metagenome-assembled

65   genomes (MAGs) of Accumulibacter have been obtained [12, 16–20] and amongst them, only

66   one complete draft genome sequence for Clade IIA, namely Accumulibacter UW1 has been

67   reported. These MAGs have highlighted the enormous genomic diversity in Accumulibacter

68   clades (>700 unique genes in each clade) especially in terms of the capacity to perform the

69   different steps of the denitrification pathway and ability to metabolize complex carbon sources

70   [16]. However, besides the MAGs from lab-scale reactors, Accumulibacter genome assemblies

71   from full-scale systems have not been reported, possibly due to their low relative abundance (< 5

72   % based on 16S rRNA gene sequencing) [10, 14] and insufficient depth of sequencing, which

73   makes MAG extraction from metagenome assemblies difficult. Metagenomic studies have

74   indicated high-level of diversity within Accumulibacter populations in full-scale EBPR

75   communities [10, 14], however this diversity was not fully characterized due to the inability to

76   extract high-quality MAGs. This phylogenetic diversity within Accumulibacter populations

77   along with the genomic functional diversity across clades points towards ecological niche-

78   partitioning due to determinants such as type of carbon source, phosphorus levels, length of

79   anaerobic period, and presence of alternate electron acceptors such as nitrate and nitrite etc.

80   Characterization of the microdiversity along with generation of corresponding MAGs could lead

81   to a better understanding of the niche determinants of Accumulibacter ecology which would help

82   better inform treatment process optimization and operation.

83   Current methods to characterize clade-level differences in Accumulibacter populations include

84   qPCR, FISH, *ppk1* gene sequencing and metagenomic sequencing. qPCR and FISH have been

85   extensively used to characterize the clade-level differences in lab-scale and full-scale systems

86   [21–24]. However, they target only those currently known clades for which primers or probes are

87   available. Metagenomic and *ppk1* gene sequencing are untargeted analyses and are capable of

88   providing phylogenetic information, however they are cost-prohibitive and/or time-consuming

89   for routine analysis at full-scale practice. Although, 16S rRNA gene amplicon sequencing has

90   emerged to be a high-throughput and widely-used method to characterize and capture broad

91   shifts in microbial communities, it is recognized that it fails to capture the phylogenetic diversity

92   within OTUs (also called microdiversity) [25, 26]. Recent studies have highlighted the

93   importance of capturing this diversity and that genetic variation in 16S rRNA gene amplicon

94   sequences can capture meaningful ecologically distinct units [27, 28]. Recent work has argued

95   for the use of a higher similarity threshold [25, 26] or for the use of exact sequence variants

96   (ESVs) or amplicon sequence variants (ASVs) [29, 30] to more accurately characterize the

97   microdiversity in a community. Oligotyping has been proposed as a method that can distinguish

98   meaningful sequence variation from sequencing errors and can partition sequence types that

99   differ in as little as one nucleotide with the assumption that the variation in small regions of the

100   16S rRNA gene translate to phylogenetic differences [27, 28, 30].

101   In this study, we employed 16S rRNA gene sequencing, oligotyping and metagenomic analysis

102   to elucidate and compare the microbial ecology and potential functional traits of Accumulibacter

103   between conventional and S2EBPR process configurations in full-scale systems. The unique

104   opportunity to operate two different EBPR configurations, in parallel and separate treatment

105   trains with the same influent feed, made it possible to eliminate other influent characteristics-

106   related selection factors and, therefore enabled association of specific CAP clades with S2EBPR

107   versus conventional A2O configuration. To overcome the limitation of 16S rRNA gene

108   sequencing and OTU-based approaches in revealing microdiversity, we demonstrated and

109   validated the use of oligotyping in revealing the microdiversity within Accumulibacter

110   populations. The results of the oligotyping analysis were further compared and confirmed with

111   phylogenetic analysis of both full-length 16S rRNA and *ppk1* gene sequences retrieved from

112   metagenomic sequencing. In addition, using differential coverage binning, we successfully

113   assembled and binned 2 high-quality and 1 medium-quality [31] Accumulibacter MAGs from

114   full-scale EBPR sludge for the first time, including a distinct *Ca.* Accumulibacter clade that is

115   dominant and associated with the S2EBPR configuration. Lastly, comparative genomic analysis

116   was performed with the extracted MAGs and with previously published MAGs, to determine

117   differences in key metabolic pathways and functions including unique functions that could

118   contribute to niche-partitioning in the dynamic and complex environments in full-scale EBPR

119   facilities.

Materials and Methods

Rock Creek Pilot

122   Full-scale testing was conducted from June 22nd to August 30th 2016 at the Rock Creek Facility

123   (Hillsboro, Oregon). Two parallel trains with separated secondary clarifiers were operated with

124   the same influent wastewater side-by-side, with one operated as conventional EBPR and another

125   one as S2EBPR configuration for comparison. The S2EBPR configuration implemented was

126   side-stream RAS fermentation with supplemental carbon addition (SSRC). In this configuration,

127   100% of the RAS was diverted to the side-stream reactor (Figure S1) for a total anaerobic HRT

128   of 1.5 hours with primary sludge fermentate addition. The side-stream reactor was mixed

129   intermittently once every week for 10 minutes during the sampling period for this study. The

130   conventional EBPR configuration was an A2O process with an anaerobic HRT of 0.7 hours.

131   Average influent flow to each treatment trains during the testing period was 5.35 MGD.

132   Throughout the testing period, process performance parameters were routinely monitored.

133   Detailed information on operation and performance is available in supplementary methods and

134   published literature [4, 6, 32].

135   *Sampling, DNA Extraction, Amplicon Sequencing and Analysis*

136   Sampling was performed before pilot testing was started as a baseline (April 7th and April 26th,

137   2016) and during pilot testing (June 22nd, July 20th, August 2nd, August 17th, and August 30th,

138   2016). MLSS samples were collected at the end of the aerobic zone, shipped to Northeastern

139   University, Boston MA overnight on dry ice. DNA was extracted according to the MiDAS

140   protocol [33]. The extracted DNA was sent to University of Connecticut-MARS facility for PCR

141   amplification and sequencing targeting the V4 region using the primers 515F (5'-

142   GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') [34]

143   and the amplicons were sequenced on the Illumina MiSeq using V2 chemistry using paired-end

144   (2 X 250) sequencing [35]. Raw reads have been submitted to NCBI under the BioProject

145   accession number PRJNA530271.

146   The sequences were trimmed to remove primers and barcodes, quality filtered using sickle v1.33

147   [36] with a minimum quality score of 20 and analyzed as described in Kozich et al., [35].

148   Consensus taxonomy of OTUs was determined using the 80% cugtoff using the MiDAS (v123)

149   database [33] (details of the analysis are provided in the supplementary methods). Amplicon data

150   was rarefied to the minimum total sequence count (12090 sequences) across all samples. All data

151   and statistical analysis were performed in R (for more details see supplementary methods) using

152   the following packages: vegan (Oksanen et al. 2007), ggplot2 [37], dplyr [38] and ampvis [39].

153   *Oligotyping Analysis*

154   To generate the input sequences required for oligotyping analysis, the Kozich et al., [35]

155   pipeline, described above, was executed without the preclustering step as recommended by Eren

156   et al. [30]. After chimera removal and denoising sequences, the 16S rRNA gene amplicon (V4

157   region) sequences classified as *Ca.* Accumulibacter were extracted and formatted as required by

158   the oligotyping pipeline using the mothur2oligo.sh script

159   (https://github.com/michberr/MicrobeMiseq/tree/master/mothur2oligo). A total of 38855

160   sequences that were classified as *Ca.* Accumulibacter were extracted and used as input for

161   oligotyping analysis. The oligotyping pipeline (v2.1) was used according to Eren et al. [30] with

162   recommended best practices. Quality filtering in the oligotyping pipeline resulted in 37934 reads

163   (97.6 %). The Shannon entropy at each nucleotide position was calculated using the entropy-

164   analysis command. Starting with the highest entropy positions, 9 nucleotide positions (56, 57, 63,

165   75, 112, 113, 114, 115, 123) were selected for entropy decomposition (Figure S2). Noise filtering

166   was performed by limiting our analysis to oligotypes that occurred in at least 3 samples (-s) and

167   those that had a minimum count (-M) of 30 (~0.1 %). A total of 195 raw oligotypes were initially

168   identified based on the 9 nucleotide positions. Denoising and elimination based on the -s and -M

169   parameters resulted in 9 oligotypes.

170   ***Metagenomic Sequencing, Assembly and Genome Binning***

171   Samples from June 26th, August 2nd and August 30th, 2016 from both treatment trains were sent

172   to the TUCF Genomics center for 2 x 250 paired end sequencing on the Illumina HiSeq 2500.

173   Library preparation for metagenomic sequencing was performed using the TruSeq PCR-free

174   DNA kit. A total of 83.2 million paired-end reads were obtained. Sequences were quality filtered

175   using sickle v1.33 [36] with a quality threshold of 20 and a length threshold of 50. In order to

176   remove contamination sequences, the quality filtered sequences were then mapped to the UniVec

177   database [40] and mapped reads were removed. Paired-end reads for all 6 samples were then co-

178   assembled together using MEGAHIT [41] using a minimum contig length of 1000 to enable

179   recovery of MAGs and eliminate erroneous contigs. Samples (n=3) from each treatment train

180   were also co-assembled separately. Quality for the assemblies (total 3 assemblies) was assessed

181   using QUAST [42]. The paired-end reads obtained after Univec-filtering were mapped to the

182   assemblies using Bowtie 2.0 [43]. The resulting sam files were converted to bam, sorted and

183   indexed using samtools [44]. Duplicate reads were removed using Picard [45]. Assembled

184   contigs were then classified using Kaiju [46] with the default database. Genome binning was

185   performed using the CONCOCT [47] implementation in Anvi'o [48] with a cluster size of 350.

186   Resulting bins with >70 % completion were then manually refined in Anvi'o based on coverage

187   patterns, GC content and taxonomic affiliation. The quality of all refined bins was then assessed

188   with CheckM v1.0.11 and taxonomic classification was also obtained. Genome bins from the

189   different assemblies were dereplicated using dRep [49] and the dereplicated bins were then used

190   for downstream analysis. Bins, classified as *Rhodocyclaceae*, were extracted and paired-end

191   reads from each sample were mapped against the MAGs using bbmap (v37.78) with 90%

192   minimum identity while choosing only the best hit and matching paired-end reads. Contig-wise

193   coverage was calculated using the pileup.sh script

194   (https://github.com/BioInfoTools/BBMap/blob/master/sh/pileup.sh). Relative abundance of

195   genome bins was calculated using the following equation (assuming an average genome length

196   of 6 Mbp):

$$Relative\ abundance = \frac{\dfrac{No.\,of\ Reads\ Mapped_{Bin}}{Length\ Mapped_{Bin}}}{\dfrac{No.\,of\ Reads\ Mapped_{Assembly}}{Average\ Genome\ Length}}$$

197   Raw reads and identified Accumulibacter MAG assemblies have been submitted to NCBI and

198   are accessible under the BioProject accession number PRJNA530189.

199 *Reconstruction of full-length 16S rRNA Gene Sequences*

200 EMIRGE [50] was used to reconstruct 16S rRNA gene sequences from the paired-end

201 metagenomic data. Exact sequence matches to oligotypes were extracted, aligned with the

202 reference sequences from He et al., [21] and a phylogenetic analysis was performed using

203 MrBayes 3.2 [51] (see supplementary methods for details).

204 *ppk1 Annotation and Phylogenetic Analysiss*

205 *ppk1* genes were annotated in the assemblies using a HMM model [52] using HMMER 3.1b2

206 (http://hmmer.org/). The annotated genes were then aligned with the reference sequences

207 (n=1012) using MUSCLE v3.8.31 [53] and placed on a reference backbone tree using pplacer

208 [54]. Read counts for each of the identified ppk1 sequences were obtained using the multicov

209 tool in BEDtools [55] and the reads per kilobase million (RPKM) value was calculated.

210 Sequences with length less than 600 bp were filtered out and relative abundance of the *ppk1* gene

211 sequences was calculated using the following equation:

$$relative\ abundance = \frac{RPKM_{ppk1\ sequence}}{sum(RPKM)_{sample}}$$

212 *Pangenomic and Phylogenetic Analysis*

213 All publicly available genomes for *Ca.* Accumulibacter (n=19) were downloaded and combined

214 with the CAP MAGs obtained in this study to perform a pangenomic analysis using the Anvi'o

215 5.1 [48] pangenomic workflow. Functional annotation of the MAGs was performed using

216 GhostKOALA [56]. A set of 37 marker genes were identified using Phylosift and a concatenated

217 alignment of these were used to construct a maximum-likelihood (ML) tree in RaxML version

218 8.2.11 [57] with a 100 bootstraps and with GTRGAMMA settings. For the phylogenetic analysis,

219 published *Propionivibrioeu* (*Propionivibrio dixarboxylicus*, *Propionivibrio militaris*, *Ca.*

220   Propionivibrio aalborgensis) and *Dechloromonas aromatica* RCB genomes were used as

221   outgroups. The best-scoring ML tree was used. The *ppk1* gene was also annotated in all the

222   genomes and used for a phylogenetic analysis using pplacer as outlined previously.

223   **Results and Discussion**

224   ***No observed differences in Accumulibacter diversity and relative abundance were observed at***

225   ***genus or OTU-level between conventional A2O and S2EBPR SSRC configurations***

226   In this study, Accumulibacter was found to be the most abundant PAO in both EBPR

227   configurations (Figure S2) [4, 6]. The estimated relative abundance of Accumulibacter, based on

228   16S rRNA gene sequencing, increased from the start of the pilot (June 21st, 2016) until the end

229   (August 30th, 2016) in both treatment trains from an average abundance of 0.35 ± 0.05 % (both

230   conventional and S2EBPR) to 1.98 ± 0.12 % (both conventional and S2EBPR) [4, 6]. A genus-

231   level analysis did not reveal significant differences between the two EBPR configurations in the

232   diversity and relative abundance of putative PAOs such as Accumulibacter. At the OTU-level,

233   both treatment trains showed a predominance of OTU 00015 with most of the total

234   Accumulibacter relative abundance accounted for by OTU 00015 (Figure 1). The rest of the

235   OTUs were present at very low relative abundances. The results suggested that the diversity in

236   Accumulibacter populations, if any, could not be resolved using OTU-based methods. However,

237   there is still potential for finer-resolution diversity which could be masked within the OTUs.

238   ***Oligotyping reveals differences in Accumulibacter clades between S2EBPR SSRC and***

239   ***conventional A2O configurations***

240   In order to further resolve the microdiversity within Accumulibacter OTUs, oligotyping analysis

241   was performed using the individual 16S rRNA gene sequences targeting the V4 region. The

242   oligotyping analysis enabled the identification of 9 oligotypes of Accumulibacter and therefore

243   distinct Accumulibacter community structure between the two EBPR configurations. The

244   S2EBPR Accumulibacter community became increasingly dominated by Oligotype 2, during the

245   testing period, while it remained at a very low relative abundance in the conventional A2O

246   configuration (Figure 2A, Table S1). The relative abundance of Oligotype 2 increased from 31.8

247    $\pm$ 8.1 % at the start of testing (Jun 22) to 68.0 $\pm$ 0.2 % on Aug 17th and 62.4 $\pm$ 1.3 % on Aug

248   30th in S2EBPR while it remained below 20 % of the Accumulibacter sequences in the parallel

249   conventional A2O treatment trains. The continuously increasing abundance of the Oligotype 2 in

250   S2EBPR over a time-series and its consistently low abundance in A2O suggests that this

251   oligotype was most likely favored and enriched by conditions in the S2EBPR configuration. In

252   comparison, predominant oligotypes in A2O were Oligotypes 5 and 6 (Figure 2A). A

253   hierarchical clustering analysis of the samples based on the oligotyping data showed distinct

254   clusters based on operating regime (Figure S4). All samples from the S2EBPR train clustered

255   together except for the sample from June 22nd which was the date when the piloting was started.

256   Samples from A2O taken during the later stages of piloting (Aug 17th, Aug 30th) clustered

257   together while the ones during the earlier stages (Jun 22nd, Jul 20th, and Aug 2nd) clustered

258   separately.

259   ***Phylogenetic analysis of oligotype sequences shows congruency with ppk1-phylogeny of***

260   ***Accumulibacter***

261   In order to further correlate the oligotypes to clade/species level differences in CAP and examine

262   the congruency of the oligotype phylogeny with *ppk1* phylogeny, we performed phylogenetic

263   analysis using near full-length 16S rRNA gene and *ppk1* sequences through metagenomic

264   sequencing on the samples taken on June 22nd, August 2nd, and August 30th from both

265   treatment trains. A phylogenetic analysis of near full-length 16S rRNA gene sequences

266   (extracted from metagenomic sequencing), classified as Accumulibacter and matching 5 of 9

267   oligotypes, along with reference sequences from He et al., [21], which were previously assigned

268   to clades based on congruency with *ppk1* phylogeny, revealed that the 16S rRNA gene sequences

269   matching the oligotypes were monophyletic (Figure 2B). It is important to note that multiple full-

270   length 16S rRNA gene sequences with identical V4 regions matching each oligotype were

271   obtained and they all clustered within the same clade. The sequences matching Oligotype 2,

272   which is the dominant oligotype in S2EBPR, clustered with Clade IIC/D 16S rRNA gene

273   sequences (He et al., 2007). Oligotype 6 clustered with Clade Clade IIA 16S rRNA gene

274   sequences. Oligotype 5 did not cluster with any representative 16S rRNA reference sequence at a

275   sub-clade level, but clustered with Clade II sequences. This could be due to the lack of a

276   representative reference sequences for that particular sub-clade.

277   Sub-clades within Accumulibacter has been previously resolved using the *ppk1* gene [17, 58–

278   61]. A phylogenetic analysis of *ppk1* gene sequences, extracted from metagenomic sequencing

279   and assigned clades using pplacer (Figure S5), was performed and the cumulative relative

280   abundance of sequences belonging to each clade was calculated (Figure 2C). This further

281   confirmed that Clade IIC was predominant in S2EBPR which is congruent with oligotyping

282   results. However, according to *ppk1* phylogeny, Clade IIG was the predominant clade in the

283   conventional A2O configuration at the end of testing which was not directly congruent with the

284   results using 16S rRNA oligotyping phylogeny. This could be due to the lack of representative

285   16S rRNA gene reference sequences for Clade IIG.

286   Previous research has cautioned that exact sequence variants (ESVs) or oligotypes may not

287   represent phylogenetically or ecologically significant units [62]. However, this was based on a

288   study of *Microcystis* in freshwater systems where they observed that the oligotypes observed in

289   the study were not monophyletic and did not correlate with toxin producing capability. In this

290   study, we show that the dominant oligotypes observed are distinct between EBPR configurations

291   fed with the same influent wastewater and their full-length 16S rRNA gene sequence matches

292   from metagenomic reconstruction are monophyletic. Furthermore, we show that the oligotyping

293   results agree with the phylogenetic analysis with a more robust marker gene (*ppk1*) for

294   Accumulibacter. It should be noted that, until now, the only way to resolve clade-level

295   differences in Accumulibacter has been using the *ppk1* gene or using metagenomics. Even

296   though qPCR primers for the *ppk1* gene exist for many clades, it is unclear whether these capture

297   the diversity of Accumulibacter sufficiently. The potential advantages of resolving clade-level

298   differences using a high-throughput method such as 16S rRNA gene amplicon sequencing

299   combined with oligotyping is significant since it will enable more cost-effective and improved

300   characterization of EBPR communities. However, the use of this method needs to be further

301   validated by comparing with other tools that can differentiate closely-related taxa in order to

302   characterize the microdiversity in a particular taxon.

303   ***Accumulibacter MAGs binned from the metagenomic sequences reveal previously***

304   ***uncharacterized diversity in full-scale EBPR systems***

305   To further characterize the identity and metabolic capabilities of Accumulibacter present in the

306   two full-scale EBPR systems with different configurations, we used genome-resolved

307   metagenomics to recover 3 MAGs (RC14, RC18, and RCAB4-2) classified putatively as

308   Accumulibacter. The taxonomic affiliation of these MAGs was further delineated by performing

309   a phylogenomic analysis along with 19 published Accumulibacter MAGs [13, 17, 18, 20, 63]

310   using a concatenated alignment of 37 single-copy genes (SCG). Based on the SCG phylogeny

311   (Figure s3A), the three MAGs were confirmed to be affiliated with Accumulibacter. RC14 and

312 RC18 were high-quality MAGs with a completion of 99 % and 98.2 % respectively with low

313 redundancy values of 0.2 % and 0.0 % respectively. The other MAG (RCAB4-2) had a

314 completion of 77.2% and 3.1% redundancy.

315 To further delineate the clade-level taxonomy of the MAGs, the *ppk1* gene sequence from each

316 MAG was extracted and placed on the reference tree as described previously. Based on their

317 placement (Figure s3B), MAG RC14 can be classified as Clade IIC and RCAB4-2 as Clade IIB.

318 Based on both SCG and *ppk1* phylogeny, MAG RC18 does not cluster under any of the

319 canonical groupings at the sub-clade level, which suggests that it is a previously uncharacterized

320 sub-clade of Accumulibacter since MAGs from full-scale facilities have not been published

321 before. Both the SCG and *ppk1* phylogeny, suggest that RC18 would belong to Clade II.

322 Average nucleotide identity (ANI) was used to confirm the results of the phylogenetic analysis

323 (Figure S7). MAG RC14 had a > 90% ANI with Clade IIC MAGs over an alignment fraction of

324 57.0 % (BA91), 75.3 % (HKU2), 62 % (SK01), 70.8 % (SK02) and 72 % (UBA5574). This

325 would confirm the classification of RC14 as a Clade IIC genome [64]. MAG RC18 and RCAB4-

326 2 had an alignment fraction of < 10 % and < 20 % respectively with any of the other MAGs with

327 ANIs of < 85 % again reinforcing the results of the phylogenetic analysis that these are

328 previously uncharacterized clades [64]. We also calculated the relative abundance of these

329 MAGs in the metagenome and observed that RC14 (Clade IIC) was the predominant

330 Accumulibacter MAG in the S2EBPR metagenome (Figure 3C). Note that the clade affiliation of

331 these genome bins corresponds to the predominant clades observed using both oligotyping and

332 *ppk1* annotation (Figure 2). Clade IIC MAG (RC14) was the most abundant Accumulibacter

333 MAG in S2EBPR, thus, again, confirming the oligotyping approach for resolving the

334 microdiversity within CAP phylogeny.

335   The newly identified RC18 genome seems to belong to a previously uncharacterized clade while

336   RCAB4-2 is classified as Clade IIB for which there have not been any representative MAGs

337   published. Our results revealed and highlighted the as-yet uncharacterized diversity of

338   Accumulibacter. Particularly, draft genomes from full-scale systems have previously not been

339   published, therefore the MAGs obtained in this study could expand the current knowledge on the

340   genomic diversity in Accumulibacter in full-scale EBPR systems.

341   ***Comparative Genomics of Ca. Accumulibacter***

342   A comparative genomic analysis of 3 CAP MAGs obtained in this study along with 19 published

343   CAP MAGs was performed with a focus on pathways relevant to EBPR metabolism such as

344   polyP metabolism, phosphate transporters, carbon cycling, and denitrification (Figure 4).

345   *Core EBPR Metabolism*

346   The characteristic metabolism of PAOs in EBPR is the concurrent cycling of carbon and

347   phosphorus through storage and use of intracellular polymers such as glycogen,

348   polyhydroxyalkanoates (PHA) and polyphosphate (polyP). The MAGs extracted in this study

349   (RC14, RC18 and RCAB4-2) were annotated with all genes (Figure 4 & Table S3) considered

350   essential for phosphate transport and polyP metabolism, including *pstSCAB, pitA, ppk1*, *ppk2* and

351   *ppx*. It has been postulated that the presence of the low-affinity Pit transporter is vital to the

352   PAO-phenotype [65]. The concerted action of these enzymes has been shown to be essential for

353   the PAO phenotype and are present in all Accumulibacter MAGs including the three identified in

354   this study [66, 67]. Genes for PHA synthesis, glycolysis and TCA cycle from both acetyl-CoA

355   and propionyl-CoA were shared among all the MAGs. One of the debated features of the carbon

356   metabolism in Accumulibacter is whether anaerobic glycolysis proceeds via the ED or the EMP

357   pathway [68]. The consistent presence of the EMP pathway and the absence of genes for the ED

358 pathway in the currently available 22 MAGs of Accumulibacter, including 3 MAGs from full-

359 scale facilities in this study, suggests that the EMP pathway is likely the predominant

360 metabolism for anaerobic glycolysis (Figure 4 & Table S3). Another uncertainity in the

361 anaerobic metabolism of EBPR is the source of the reducing equivalents for PHA formation

362 under anaerobic conditions. Current hypothesized pathways include glycolysis, anaerobic

363 operation of the TCA cycle, glyoxylate shunt and split TCA cycle [69–71]. Key enzymes for all

364 these proposed pathways are present in the 3 MAGs from this study (Figure 4). This along with

365 trancriptomic evidence from literature [11, 15, 18, 72, 73] suggests that Accumulibacter likely

366 possess phenotypic versatility and can utilize various pathways in various combinations,

367 depending on the availability of external substrate and intracellular polymers, for their anaerobic

368 metabolism [71, 74, 75].

369 *Carbon Metabolism*

370 One of the critical aspects of PAO-type metabolism is the uptake of low molecular weight

371 compounds such as VFAs (primarily acetate and propionate) under anaerobic conditions to form

372 PHA. The 3 Accumulibacter MAGs, obtained in this study, contain the genes necessary for

373 passive transport of acetate and propionate (*actP*) and conversion of acetate and propionate into

374 acetyl-CoA and propionyl-CoA through acetate-coA synthetase (*acsA*) and propionate-coA

375 ligase (*prpE*) respectively (Figure 4 & Table S3).

376 Full-scale facilities contain a variety of substrates in the wastewater including complex

377 substrates, therefore the ability to use complex substrates can provide a niche for a PAO in full-

378 scale facilities. Most of the MAGs of Accumulibacter, including the 3 MAGs from this study,

379 encode a lactate dehydrogenase gene (*dld*) (except for BA94 and SK01) which catalyzes the

380 conversion of lactate to pyruvate and the reverse reaction. The lactate permease gene (*lctP*/lldP)

381   was identified in aalborgensis, BA91, HKU2, UW1, RC18 and RCAB4-2. The presence of the

382   lactate permease gene, in combination with lactate dehydrogenase, could confer the ability to

383   uptake lactate and convert into pyruvate, acetyl-CoA or propionyl-CoA which can then be used

384   to form PHB/PHV or feed the TCA cycle. Conversion of lactate to pyruvate is able to generate

385   electron equivalents which can also be used for other active transport processes. A recent study

386   [76] showed that Accumulibacter is capable of using lactate to store PHA, however, the authors

387   noticed a deterioration in EBPR performance when feeding lactate presumably associated with a

388   metabolic shift in Accumulibacter from utilizing both polyP and glycogen to utilizing just

389   glycogen (similar to GAOs) since the use of lactate doesn't require the hydrolysis of polyP to

390   generate ATP.

391   *Nitrogen Metabolism*

392   Another trait of interest in Accumulibacter metabolism is the ability to use various oxidized

393   nitrogen species as electron acceptors. Among the available MAGs, only those in Clade IIC

394   MAGs, including MAG RC14 that was dominant in the S2EBPR configuration in this study and

395   one exception from IIA (aalborgensis), contain gene sets for respiratory nitrate reduction

396   (*narGHI*) and nitrite reduction (*nirS*), but did not possess genes for subsequent steps of

397   denitrification. The only MAG that contains genes for a complete denitrification pathway is UW-

398   LDO-IC (Clade IC). Recent work with metatranscriptomics [20] showed that under micro-

399   aerobic conditions and in the presence of nitrate, UW-LDO-IC is able to use both nitrate and

400   oxygen simultaneously as electron acceptors even though the ability to use nitrate as an electron

401   acceptor with exclusive and concurrent uptake of phosphate was not shown.

402   Most other clades (Clade IIA, IIB, IIF, IA, and IB) contain genes for periplasmic dissimilatory

403   nitrate reduction (*napAB*) system. The Nap system is generally considered to not be coupled with

404   PMF generation and anaerobic respiration. It has been suggested that this system might be

405   involved in aerobic denitrification for roles in adaptation to anaerobic metabolism after transition

406   from aerobic conditions such as in EBPR [78]. There is also evidence that Nap is a dissimilatory

407   enzyme used for optimizing redox balancing [79, 80]. The possession of periplasmic Nap system

408   enzymes in majority of the Accumulibacter clades seem to suggest that these Accumulibacter

409   likely do not use nitrate as an electron acceptor for respiration and denitrification. Rather, the

410   Nap system likely allows CAPs to aerobically reduce nitrate to nitrite in the periplasm, which

411   can be utilized for downstream nitrite reduction or excreted outside the cell.

412   It is possible that Accumulibacter species with truncated denitrification pathways could use

413   complementary pathways in flanking populations to perform complete denitrification on a

414   community level [77]. Presence of genes for nitric oxide reduction are restricted to a few MAGs

415   in Clade II (UW1, RCAB4-2, BA91) and most Clade I MAGs. In contrast, genes for nitrous

416   oxide reduction are present in almost all Clade IIF, Clade I MAGs and UW1 (Clade IIA).

417   Noticeably, RC18 (present in both the conventional and S2EBPR full-scale systems) did not

418   possess any of the reductases in the denitrification pathway, despite having a completion of 98.2

419   %.

420   *Unique Functions in MAGs from Full-Scale Systems*

421   In this study, we were able to obtain 3 Accumulibacter MAGs from full-scale EBPR

422   communities. It is likely that MAGs from full-scale facilities have some unique functional

423   abilities compared to MAGs from enriched lab-scale bioreactors to enable the organism to

424   survive in the highly dynamic and complex environmental conditions in wastewater treatment

425   systems. We identified genes for dimethyl sulfoxide (DMSO) reduction exclusively in RC18.

426   The pathway included the *dmsA* and *dmsB* subunits of the *dmsABC* gene cluster. The DMSO

427   respiratory pathway has been extensively studied in photo-heterotrophic organisms such as

428   *Rhodobacter* spp. and plays a role in the organism's redox homeostasis and enables the

429   organisms to utilize a diversity of carbon sources during dark anaerobic growth [80]. Its potential

430   role in Accumulibacter metabolism is unknown and warrants further investigation. We also

431   identified the gene for alkane oxidation, alkane 1-monooxygenase (*alkB*) in RCAB4-2. The

432   production of this enzyme enables the oxidation of alkanes to alcohols making it an important

433   enzyme for bioremediation of petroleum hydrocarbon contaminated environments [81, 82].

434   Whether this enzyme has a direct impact on EBPR metabolism is unknown.

435   In this study, for the first time, using integrated amplicon sequencing, oligotyping and genome-

436   resolved metagenomics, we were able to reveal clade-level differences in Accumulibacter

437   communities and associate the differences with two different full-scale EBPR configurations.

438   The results led to the identification and characterization of a distinct and dominant

439   Accumulibacter oligotype - Oligotype 2 (belonging to Clade IIC) and its matching MAG (RC14)

440   associated with S2EBPR configuration. This study demonstrates that oligotyping of the V4

441   region of the 16S rRNA gene could resolves clade-level differences, congruent with *ppk1*-

442   phylogeny, in Accumulibacter communities in full-scale EBPR systems. The potential

443   advantages of resolving clade-level differences using a high-throughput method such as 16S

444   rRNA gene amplicon sequencing combined with oligotyping is significant since it will enable

445   wider capture, higher-resolution and more cost-effective genotyping of Accumulibacter

446   communities and thus improve our ability to characterize and monitor EBPR processes.

447   Genome-resolved metagenomics also enabled us to retrieve 3 new CAP MAGs from full-scale

448   facilities including two MAGs, RCAB4-2 and RC18, belonging to Clade IIB and to a previously

449   uncharacterized sub-clade of Clade II respectively, that did not have any previous representative

450　genomes. Comparative genomics of the obtained MAGs showed unique functions that could

451　enable niche-partitioning of particular clades under different conditions.

Acknowledgements

Conflict of Interest

458　The authors declare no conflict of interest.

459

References

461

462　1.　Morse GK, Brett SW, Guy JA, Lester JN. Review: Phosphorus removal and recovery

463　　　technologies. *Sci Total Environ* 1998; **212**: 69–81.

464　2.　Gu AZ, Saunders A, Neethling JB, Stensel HD, Blackall LL. Functionally Relevant

465　　　Microorganisms to Enhanced Biological Phosphorus Removal Performance at Full-Scale

466　　　Wastewater Treatment Plants in the United States. *Water Environ Res* 2008; **80**: 688–698.

467　3.　Onnis-Hayden A, Srinivasan V, Tooker NB, Li G, Wang D, Gu AZ. Survey of Full Scale

468　　　Side-Stream EBPR Facilities and Comparison with Conventional EBPR: Process Stability,

469　　　Kinetics and Microbial Ecology. *Preprints* 2018.

470　4.　Gu AZ, Tooker NB, Onnis-Hayden A, Wang D, Srinivasan V, Li G, et al. Optimization

471　　　and Design of a Side-Stream EBPR Process as a Sustainable Approach for Achieving

472　　　Stable and Efficient Phosphorus Removal. 2019.

473   5.   Barnard JL, Kobylinski E. The Case for Side-Stream RAS or Mixed Liquor Fermentation

474   To Enhance Biological Phosphorus Removal (EBPR). *Proc. Water Environ. Fed. Nutr.*

475   2018. Raleigh, NC.

476   6.   Wang D, Tooker NB, Srinivasan V, Li G, Schauer P, Menniti A, et al. A Full-Scale

477   Comparative Study of Conventional and Side-Stream Enhanced Biological Phosphorus

478   Removal Processes. *Preprints* 2018.

479   7.   Mielczarek AT, Nguyen HTT, Nielsen JL, Nielsen PH. Population dynamics of bacteria

480   involved in enhanced biological phosphorus removal in Danish wastewater treatment

481   plants. *Water Res* 2013; **47**: 1529–1544.

482   8.   Stokholm-Bjerregaard M, McIlroy SJ, Nierychlo M, Karst SM, Albertsen M, Nielsen PH.

483   A Critical Assessment of the Microorganisms Proposed to be Important to Enhanced

484   Biological Phosphorus Removal in Full-Scale Wastewater Treatment Systems. *Front*

485   *Microbiol* 2017; **8**: 718.

486   9.   Flowers JJ, Cadkin TA, McMahon KD. Seasonal bacterial community dynamics in a full-

487   scale enhanced biological phosphorus removal plant. *Water Res* 2013; **47**: 7019–7031.

488   10.   Law Y, Kirkegaard RH, Cokro AA, Liu X, Arumugam K, Xie C, et al. Integrative

489   microbial community analysis reveals full-scale enhanced biological phosphorus removal

490   under tropical conditions. *Sci Rep* 2016; **6**: 25719.

491   11.   Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. Metatranscriptomic

492   insights on gene expression and regulatory controls in Candidatus Accumulibacter

493   phosphatis. *ISME J* 2016; **10**: 810–822.

494   12.   Martín HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, et al.

495   Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge

496   communities. *Nat Biotechnol* 2006; **24**: 1263–1269.

497   13.   Flowers JJ, He S, Malfatti S, del Rio TG, Tringe SG, Hugenholtz P, et al. Comparative

498   genomics of two 'Candidatus Accumulibacter' clades performing biological phosphorus

499   removal. *ISME J* 2013; **7**: 2301–2314.

500   14.   Albertsen M, Hansen LBS, Saunders AM, Nielsen PH, Nielsen KL. A metagenome of a

501   full-scale microbial community carrying out enhanced biological phosphorus removal.

502   *ISME J* 2012; **6**: 1094–1106.

503   15.   Barr JJ, Dutilh BE, Skennerton CT, Fukushima T, Hastie ML, Gorman JJ, et al.

504   Metagenomic and metaproteomic analyses of Accumulibacter phosphatis-enriched

505   floccular and granular biofilm. *Environ Microbiol* 2016; **18**: 273–287.

506   16.   Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. Expanding our view of genomic

507   diversity in *Candidatus* Accumulibacter clades. *Environ Microbiol* 2015; **17**: 1574–1585.

508   17.   Albertsen M, McIlroy SJ, Stokholm-Bjerregaard M, Karst SM, Nielsen PH. '*Candidatus*

509   Propionivibrio aalborgensis': A novel glycogen accumulating organism abundant in full-

510   scale enhanced biological phosphorus removal plants. *Front Microbiol* 2016; **7**.

511   18.   Mao Y, Yu K, Xia Y, Chao Y, Zhang T. Genome reconstruction and gene expression of '

512   candidatus accumulibacter phosphatis' Clade IB performing biological phosphorus

513   removal. *Environ Sci Technol* 2014; **48**: 10363–10371.

514   19.   Camejo PY, Oyserman BO, McMahon KD, Noguera DR. Integrated Omic Analyses

515   Provide Evidence that a " *Candidatus* Accumulibacter phosphatis" Strain Performs

516   Denitrification under Microaerobic Conditions. *mSystems* 2019; **4**: 1–23.

517   20.   Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al.

518   Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree

519   of life. *Nat Microbiol* 2017; **2**: 1533–1542.

520   21.   He S, Gall DL, McMahon KD. 'Candidatus accumulibacter' population structure in

521   enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase

522   genes. *Appl Environ Microbiol* 2007; **73**: 5865–5874.

523   22.   Camejo PY, Owen BR, Martirano J, Ma J, Kapoor V, Santo Domingo J, et al. Candidatus

524   Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic

525   conditions simultaneously use different electron acceptors. *Water Res* 2016; **102**: 125–

526   137.

527   23.   Welles L, Abbas B, Sorokin DY, Lopez-Vazquez CM, Hooijmans CM, van Loosdrecht

528   MCM, et al. Metabolic response of '*Candidatus* Accumulibacter phosphatis' clade IIC to

529   changes in influent P/C ratio. *Front Microbiol* 2017; **7**.

530   24.   Mao Y, Graham DW, Tamaki H, Zhang T. Dominant and novel clades of Candidatus

531   Accumulibacter phosphatis in 18 globally distributed full-scale wastewater treatment

532   plants. *Sci Rep* 2015; **5**: 11857.

533   25.   Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs.

534   *Bioinformatics* 2018; 1–5.

535   26.   Chase AB. The importance of resolving biogeographic patterns of microbial

536   microdiversity. *Microbiol Aust* 2018; 18–21.

537   27.   Fisher JC, Murat Eren A, Green HC, Shanks OC, Morrison HG, Vineis JH, et al.

538   Comparison of sewage and animal fecal microbiomes by using oligotyping reveals

539   potential human fecal indicators in multiple taxonomic groups. *Appl Environ Microbiol*

540   2015; **81**: 7023–7033.

541   28.   Fisher JC, Levican A, Figueras MJ, McLellan SL. Population dynamics and ecology of

542   Arcobacter in sewage. *Front Microbiol* 2014; **5**: 1–9.

543   29.   Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace

544   operational taxonomic units in marker gene data analysis. *bioRxiv* 2017; 113597.

545   30.   Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping:

546   Differentiating between closely related microbial taxa using 16S rRNA gene data.

547   *Methods Ecol Evol* 2013; **4**: 1111–1119.

548   31.   Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et

549   al. Minimum information about a single amplified genome (MISAG) and a metagenome-

550   assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.

551   32.   Maher C, Schauer P, Menniti A. Testing the Limits of EBPR for Stable Operation. *Proc.*

552   *Water Environ. Fed. WEFTEC*. 2018. Chicago, IL, pp 1581–1594.

553   33.   McIlroy SJ, Saunders AM, Albertsen M, Nierychlo M, McIlroy B, Hansen AA, et al.

554   MiDAS: The field guide to the microbes of activated sludge. *Database* 2015; **2015**: 1–8.

555   34.   Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et

556   al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.

557   *PNAS* 2011; **108**: 4516–4522.

558   35.   Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-

559   index sequencing strategy and curation pipeline for analyzing amplicon sequence data on

560   the miseq illumina sequencing platform. *Appl Environ Microbiol* 2013; **79**: 5112–5120.

561   36.   Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ

562   files (Version 1.33) [Software]. *Available at https://github.com/najoshi/sickle* . 2011. ,

563   2011

564   37.   Wickham H. ggplot2: Elegant graphics for data analysis. 2009. Springer-Verlag, New

565   York.

566   38.   Wickman H, Francois R, Henry L, Müller K. A Grammar of Data Manipulation. R

567   package version 0.7.5. 2018.

568   39.   Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. Back to basics - The

569   influence of DNA extraction and primer choice on phylogenetic analysis of activated

570   sludge communities. *PLoS One* 2015; **10**: e0132783.

571   40.   Univec database at National Center for Biotechnology Information

572   [http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html].

573   41.   Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node

574   solution for large and complex metagenomics assembly via succinct de Bruijn graph.

575   *Bioinformatics* 2015; **31**: 1674–1676.

576   42.   Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: Evaluation of metagenome

577   assemblies. *Bioinformatics* 2016; **32**: 1088–1090.

578   43.   Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;
579   **9**: 357–359.

580   44.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
581   Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

582   45.   Picard. http://broadinstitute.github.io/picard/.

583   46.   Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics
584   with Kaiju. *Nat Commun* 2016; **7**: 11257.

585   47.   Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
586   metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.

587   48.   Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an
588   advanced analysis and visualization platform for 'omics data. *PeerJ* 2015; **3**: e1319.

589   49.   Olm MR, Brown CT, Brooks B, Banfield JF. DRep: A tool for fast and accurate genomic
590   comparisons that enables improved genome recovery from metagenomes through de-
591   replication. *ISME J* 2017; **11**: 2864–2868.

592   50.   Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: Reconstruction of
593   full-length ribosomal genes from microbial community short read sequencing data.
594   *Genome Biol* 2011; **12**.

595   51.   Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed
596   models. *Bioinformatics* 2003; **19**: 1572–1574.

597   52.   Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, et al. Strain-level
598   diversity drives alternative community types in millimetre-scale granular biofilms. *Nat*

599   *Microbiol* 2018; **3**: 1295–1303.

600   53.   Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high

601   throughput. *Nucleic Acids Res* 2004; **32**: 1792–1797.

602   54.   Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and

603   Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC*

604   *Bioinformatics* 2010; **11**: 538.

605   55.   Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic

606   features. *Bioinformatics* 2010; **26**: 841–842.

607   56.   Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for

608   Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 2016;

609   **428**: 726–731.

610   57.   Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of

611   large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.

612   58.   McMahon KD, Yilmaz S, He S, Gall DL, Jenkins D, Keasling JD. Polyphosphate kinase

613   genes from full-scale activated sludge plants. *Appl Microbiol Biotechnol* 2007; **77**: 167–

614   173.

615   59.   Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. Expanding our view of genomic

616   diversity in *Candidatus* Accumulibacter clades. *Environ Microbiol* 2015; **17**: 1574–1585.

617   60.   Camejo PY, Oyserman BO, McMahon KD, Noguera DR. Integrated Omic Analyses

618   Provide Evidence that a " *Candidatus* Accumulibacter phosphatis" Strain Performs

619   Denitrification under Microaerobic Conditions. *mSystems* 2019; **4**.

620   61.   He S, Gu AZ, McMahon KD. Progress toward understanding the distribution of

621   Accumulibacter among full-scale enhanced biological phosphorus removal systems.

622   *Microb Ecol* 2008; **55**: 229–236.

623   62.   Berry MA, White JD, Davis TW, Jain S, Johengen TH, Dick GJ, et al. Are oligotypes

624   meaningful ecological and phylogenetic units? A case study of Microcystis in Freshwater

625   lakes. *Front Microbiol* 2017; **8**: 1–7.

626   63.   Skennerton CT, Barr JJ, Slater FR, Bond PL, Tyson GW. Expanding our view of genomic

627   diversity in CandidatusAccumulibacter clades. *Environ Microbiol* 2015; **17**: 1574–1585.

628   64.   Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides

629   NC, et al. Microbial species delineation using whole genome sequences. *Nucleic Acids

630   Res* 2015; **43**: 6761–6771.

631   65.   McIlroy SJ, Albertsen M, Andresen EK, Saunders AM, Kristiansen R, Stokholm-

632   Bjerregaard M, et al. 'Candidatus Competibacter'-lineage genomes retrieved from

633   metagenomes reveal functional metabolic diversity. *ISME J* 2014; **8**: 613–624.

634   66.   Kristiansen R, Nguyen HTT, Saunders AM, Nielsen JL, Wimmer R, Le VQ, et al. A

635   metabolic model for members of the genus *Tetrasphaera* involved in enhanced biological

636   phosphorus removal. *ISME J* 2013; **7**: 543–554.

637   67.   Oyserman BO, Noguera DR, del Rio TG, Tringe SG, McMahon KD. Metatranscriptomic

638   insights on gene expression and regulatory controls in Candidatus Accumulibacter

639   phosphatis. *ISME J* 2016; **10**: 810–822.

640   68.   Oehmen A, Lemos PC, Carvalho G, Yuan Z, Keller J, Blackall LL, et al. Advances in

641   enhanced biological phosphorus removal: From micro to macro scale. *Water Res* 2007;

642   **41**: 2271–2300.

643 69. Comeau Y, Hall KJ, Hancock REW, Oldham WK. Biochemical Model for Enhanced

644   Biological Phosphorus Removal. *Water Res* 1986; **20**: 1511–1521.

645 70. Mino T, Arun V, Tsuzuki Y, Matsuo T. Effect of phosphorus accumulation on acetate

646   metabolism in the biological phosphorus removal process. *Biol. Phosphate Remov. from*

647   *Wastewaters.* 1987. Pergamon, pp 27–38.

648 71. Hesselmann RPX, Von Rummell R, Resnick SM, Hany R, Zehnder AJB. Anaerobic

649   Metabolism of Bacteria Performing Enhanced Biological Phosphate Removal. *Water Res*

650   2000; **34**: 3487–3494.

651 72. He S, McMahon KD. '*Candidatus* Accumulibacter' gene expression in response to

652   dynamic EBPR conditions. *ISME J* 2011; **5**: 329–340.

653 73. Burow LC, Mabbett AN, Blackall LL. Anaerobic glyoxylate cycle activity during

654   simultaneous utilization of glycogen and acetate in uncultured Accumulibacter enriched in

655   enhanced biological phosphorus removal communities. *ISME J* 2008; **2**: 1040–1051.

656 74. Pramanik J, Trelstad PL, Schuler AJ, Jenkins D, Keasling JD. Development and validation

657   of a flux-based stoichiometric model for enhanced biological phosphorus removal

658   metabolism. *Water Res* 1999; **33**: 462–476.

659 75. Silva LG da, Gamez KO, Gomes JC, Akkermans K, Welles L, Abbas B, et al. Revealing

660   metabolic flexibility of Candidatus Accumulibacter phosphatis through redox cofactor

661   analysis and metabolic network modeling. *bioRxiv* 2018; 458331.

662   76.   Rubio-Rincón FJ, Welles L, Lopez-Vazquez CM, Abbas B, van Loosdrecht MCM,

663   Brdjanovic D. Effect of Lactate on the Microbial Community and Process Performance of

664   an EBPR System. *Front Microbiol* 2019; **10**: 1–11.

665   77.   Gao H, Mao Y, Zhao X, Liu W-T, Zhang T, Wells G. Genome-centric metagenomics

666   resolves microbial diversity and prevalent truncated denitrification pathways in a

667   denitrifying PAO-enriched bioprocess. *Water Res* 2019; **155**: 275–287.

668   78.   Moreno-Vivián C, Cabello P, Martinez-Luque M, Blasco R, Castillo F. Prokaryotic

669   Nitrate Reduction☐: Molecular Properties and Functional Distinction among Bacterial

670   Nitrate Reductases. *J Bacteriol* 1999; **181**: 6573–6584.

671   79.   Siddiqui RA, Warnecke-Eberz U, Hengsberger A, Schneider B, Kostka S, Friedrich B.

672   Structure and function of a periplasmic nitrate reductase in Alcaligenes eutrophus H16. *J

673   Bacteriol* 1993; **175**: 5867–5876.

674   80.   Richardson DJ, King GF, Kelly DJ, McEwan AG, Ferguson SJ, Jackson JB. The role of

675   auxiliary oxidants in maintaining redox balance during phototrophic growth of

676   *Rhodobacter capsulatus* on propionate or butyrate. *Arch Microbiol* 1988; **150**: 131–137.

677   81.   Smith CB, Tolar BB, Hollibaugh JT, King GM. Alkane hydroxylase gene (alkB)

678   phylotype composition and diversity in northern Gulf of Mexico bacterioplankton. *Front

679   Microbiol* 2013; **4**: 370.

680   82.   Wu B, Lan T, Lu D, Liu Z. Ecological and enzymatic responses to petroleum

681   contaminations. *Environ Sci Process Impacts* 2013; **16**: 1501–1509.

682   83.   Welles L, Tian WD, Saad S, Abbas B, Lopez-Vazquez CM, Hooijmans CM, et al.

683   Accumulibacter clades Type I and II performing kinetically different glycogen-

684   accumulating organisms metabolisms for anaerobic substrate uptake. *Water Res* 2015; **83**:
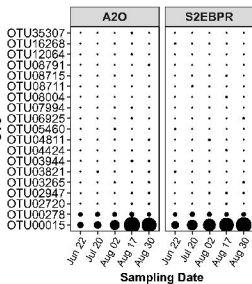
685   354–366.

686

687   **Figure Legends**

688   **Figure 1:** Comparison of most abundant OTUs classified as *Ca*. Accumulibacter in

689   Conventional A2O and S2EBPR SSRC configurations.

690   **Figure 2:** (A) Relative abundance of oligotypes of 16S rRNA gene amplicon sequences

691   classified as *Ca.* Accumulibacter in conventional A2O and S2EBPR SSRC configurations (B)

692   Phylogenetic tree of near full-length 16S rRNA gene sequences, that are exact matches to

693   oligotypes, derived from metagenomic sequences using EMIRGE (highlighted in bold) and

694   reference 16S rRNA gene sequences from He et al., (2007). Clade assignments for reference

695   sequences are according to He et al., (2007) which were based on congruency with *ppk1*

696   phylogeny (C) Relative Abundance of *ppk1* gene sequences extracted from the individual

697   treatment-train assemblies and assigned clade-classification using pplacer.

698

699   **Figure 3:** (A)Phylogeny of Ca. Accumulibcter MAGs recovered from this study along with

700   published draft genomes (n=19) of Accumulibacter, *Propionivibrio* and *Dechloromonas*

701   *aromatica* as an outgroup based on a concatenated alignment of 37 marker genes. Genome bins

702   recovered in this study are in bold. (B) *ppk1* genes from Ca. Accumulibacter genomes recovered

703   from this study placed on a backbone reference tree created with 1012 sequences and classified

704   according to He et al.,[21] (C) Relative abundance of MAGs recovered in this study.

705   **Figure 4:** Inventory of pathways in Accumulibacter MAGs associated with EBPR metabolism

706   and denitrification. Blue and grey rectangles represent presence and absence of pathways based

707   on annotation using KEGG.

708

(A)

(B)

(C)

(A)

0.05

SK02 IIC
HKU2 IIC
UBA5574 IIC
SK01 IIC
BA91 IIC
RC14
RCAB4-2
SK12 IIF
UBA6585 IIF
UBA2315 IIF
BA94 IIF
SK11 IIF
UBA2327 IIF
UW1 IIA
aalborgensis IIA
RC18
BA92 IB
UBA2783 IB
HKU1 IB
UW_LDO IC
UBA6658 IA
BA93 IA
Propionivibrio dicarboxylicus
Propionivibrio pelophilus
Ca. Propionivibrio aalborgensis
Propionivibrio militaris
Dechloromonas aromatica

Accumulibacter

(B)

Accumulibacter

Clade IID

Clade IIF

Clade II-I

Clade IIG

Clade IIC

RC14

RCAB4-2

Clade IIB

Clade IIE

Clade IIA

RC18

Clade IA

Clade IB

Clade IC

Clade ID

Clade IE

0.07

(C)

Relative Abundance (%)

A2O	S2EBPR

RC14
RC18
RCAB4-2

Jun 22   Aug 02   Aug 30   Jun 22   Aug 02   Aug 30

Sample

Clade

IA	IB	IC	II
IIA	IIB	IIC	IIF

Rows (metabolic functions):

- Cytochrome c-oxidation-cbb3
- Cytochrome c-oxidation-aa3
- High-affinity phosphate transport (PST)
- Low-affinity phosphate transport (PiT)
- PolyP Metabolism
- TCA Cycle
- EMP Pathway
- PHA Metabolism
- ED Pathway
- Split TCA Cycle
- Glyoxylate Shunt
- Acetate Metabolism
- Propionate Metabolism
- Nitrate/Nitrite Transport
- Respiratory Nitrate Reduction
- Periplasmic Nitrate Reduction
- Nitrite Reduction
- Nitric Oxide Reduction
- Nitrous Oxide Reduction

Columns (taxa): BA93, UBA6658, BA92, HKU1, UBA2783, UW-LDO-IC, RC18, aalborgensis, UW1, RCAB4-2, BA91, HKU2, RC14, SK01, SK02, UBA5574, BA94, SK11, SK12, UBA2315, UBA2327, UBA6585