**Supplementary Information for**

**Integrated Omics Analyses Reveal Differential Gene Expression and Potential for Cooperation Between Denitrifying Polyphosphate and Glycogen Accumulating Organisms**

Yubo Wang1#, Han Gao1#, George Wells1\*

1 Northwestern University, 2145 Sheridan Road, Evanston, IL, USA.

# These authors contributed equally

\*Email: [george.wells@northwestern.edu](mailto:george.wells@northwestern.edu) Phone: +1-847-491-8794

**Supplementary Information – Results**

*Carbon and Energy Metabolism and Potential Metabolite Exchange among PAOs and Flanking*

*Populations*

Figure S3 shows the expression profiles of genes involved in the major carbon and energy metabolism across the six transcriptionally most active MAGs. Apart from their key functional role in polyP accumulation, as expected, genes involved in both the glycogen metabolism and the TCA cycle were annotated as being actively expressed in the three Accumulibacter populations of clade IF, IC and IA. In addition, all three Accumulibacterpopulations, especially Acc-IF, were identified as being dominant in the expression of genes involved in EPS synthesis (Figure S3), in type IV pilus biosynthesis (*PilQ* and *PilY1*), in the biosynthesis of co-factors (e.g., biotin) and in the uptake of vitamin B12 (Figure S4).

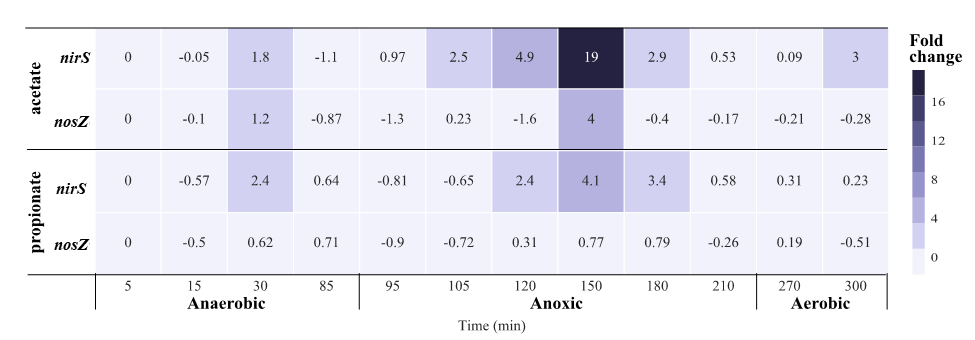
Among the three transcriptionally active non-PAO flanking populations,GAO1 likely utilizes acetate and propionate as its primary carbon source; while the activity for the utilization of the primary carbon source (acetate and propionate) is limited for both CH7 and PR6, as indicated by the expression levels of the *acs*, *ackA* and *prpE* genes. CH7 is apparently incapable of utilizing propionate, since no *prpE* gene was identified in its genome. CH7 and PR6 may probably scavenge extracellular substances (endogenous organic carbon) excreted by the PAO populations as their primary carbon source. As is summarized in Figure S4, ~88% of the *vpr* gene transcripts for extracellular protease were mapped to CH7, and ~84% gene transcripts for extracellular serine protease were associated with PR6. Both PR6 and CH7 expressed genes encoding glycoside hydrolases (xylosidase, lysozyme, glucosidase, and isoamylase) and extracellular cellulose binding enzymes that may be linked to the breakdown of polysaccharides such as xylan, starch and peptidoglycan (Figure S5). In addition, the vitamin B12 transporter *btuB* gene and *tonB* gene were also highly expressed in PR6 (Figure S4); *tonB* was reported as being essential in importing essential micronutrients across the outer cell membrane (Shultis et al., 2006). Another set of genes that were transcriptionally highly active in PR6 was the *PilQ* and *PilY1*, which two are involved in the type IV pili biosynthesis. The role of type IV pili in promoting biofilm formation has been reported in archaea ([Mechthild Pohlschroder](https://www.ncbi.nlm.nih.gov/pubmed/?term=Pohlschroder%20M%5BAuthor%5D&cauthor=true&cauthor_uid=25852657) and [Rianne N. Esquivel](https://www.ncbi.nlm.nih.gov/pubmed/?term=Esquivel%20RN%5BAuthor%5D&cauthor=true&cauthor_uid=25852657), 2015) and in *Clostridium* (Maldarelli et al., 2016). As ~33% of the *PilQ* and *PilY1* gene transcripts were expressed by the PR6, we speculate that PR6 together with the Accumulibacter PAO populations may play functional roles in this community promoting the granule aggregation (biofilm formation). This hypothesis awaits further verification.

**Supplementary Information – Tables**

**Table S1**. ANI between the draft metagenome assembled genomes from this study and the corresponding most closely related (based on ANI) publicly available reference genome (NCBI accession numbers provided in the table).

|  |  |  |
| --- | --- | --- |
| Bin\_ID | Reference genome | ANI |
| Acc-IF | GCA\_003538495.1 | 83 |
| GAO1 | GCA\_002391525.1 | 78 |
| PR6 | GCF\_000972865.1 | 87 |
| CH7 | GCA\_002746795.1 | 77 |
| BA1 | GCA\_001897835.1 | 76 |
| CH1 | GCA\_001567085.1 | 76 |
| CH6 | GCA\_002699125.1 | 76 |
| CH2 | GCA\_002352035.1 | 76 |
| BA12 | GCF\_002283555.1 | 76 |
| BA2 | GCA\_002352145.1 | 76 |
| VE2 | GCA\_002396485.1 | 76 |
| CH3 | GCA\_002842085.1 | 76 |
| MY1 | GCF\_000331735.1 | 77 |
| VE1 | GCA\_002396485.1 | 77 |
| VE3 | GCF\_000972765.1 | 77 |
| PL1 | GCA\_002343325.1 | 78 |
| BA6 | GCA\_002344125.1 | 78 |
| CH5 | GCA\_003451595.1 | 79 |
| PR2 | GCA\_002280405.1 | 79 |
| PR4 | GCF\_001293525.1 | 79 |
| PR3 | GCA\_001770955.1 | 79 |
| CH4 | GCA\_001567085.1 | 79 |
| VE4 | GCA\_002344865.1 | 80 |
| PR5 | GCF\_900155935.1 | 84 |
| BA10 | GCA\_002839825.1 | 86 |
| BA7 | GCA\_003523425.1 | 96 |
| BA8 | GCA\_002344125.1 | 98 |
| BA3 | GCA\_002344975.1 | 98 |
| BA5 | GCA\_002426785.1 | 98 |
| PR1 | GCA\_002425415.1 | 99 |
| BA11 | GCA\_001567175.1 | 100 |

**Table S2**. RT-qPCR quantification on the expression of the *nirS* and *nosZ* gene during typical SBR batches with either acetate or propionate as the primary carbon source. Fold change is relative to the 5 minute time point during the anaerobic period. Based on RT-qPCR results shown here, time points 30 minutes (anaerobic), 150 minutes (anoxic), and 270 minutes (aerobic) were selected for metatranscriptomic sequencing.



**Table S3.** Transcriptional activities (represented by RNA-RPKM values) of the functional genes annotated MAG GAO1, a putative glycogen accumulating organism affiliated with Competibacteraceae**.** Bolded rows highlight transcriptional activity of genes encoding NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (involved in anaerobic glycogen hydrolysis) and phosphoenolpyruvate carboxykinase and pyruvate phosphate dikinase (involved in glycogen synthesis).

**Table S4.** Detailed information on current publicly available Accumulibacter and Competibacteraceae reference genomes, including the ANI between GAO1 and the Competibacteraceae reference genomes, and the ANI between Acc-IF and the Accumulibacter reference genomes.

**Table S5.** Transcriptional activities (represented by RNA-RPKM values) of key functional genes annotated in all the 32 high-quality MAGs recovered in this study.

**Table S6**. Summary statistics for metagenomic sequencing and assembly

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total reads (Gbp) | Filtered Reads (Gbp) | kmer | Total assembled reads (Gbp) | Longest contig | # of Contigs | N50\* (Contigs) |
| DNA-Total\* | 9.1 | 8.6 | 70 | 0.36 | 331,655 | 165,316 | 3,707 |
| DNA-L-350 | 27.1 | 26.3 | 70 | 0.26 | 209,078 | 129,208 | 2,905 |
| DNA-S-350 | 15.5 | 14.5 | 70 | 0.43 | 331,566 | 174,811 | 4,643 |
| co-assembly | 51.7 | 49.4 | 70 | 0.66 | 412,072 | 290,166 | 4,145 |

DNA-Total: Bulk sludge DNA extracts

DNA-L-350: DNA extracts subjected to metagenomic sequencing from sludge particulates with diameter > 350 µm

DNA-S-350: DNA extracts subjected to metagenomic sequencing from sludge particulates with diameter < 350 µM.

**Table S7.** Detailed information on the denitrification genes annotated in each of the 32 high-quality MAGs.

**Table S8.** Genome accession numbers and the biosample accession numbers of all the 32 high-quality MAGs submitted to NCBI.

**Supplementary Information – Figures**

**Figure S1**. Relative abundance of the five *Accumulibacter* clades in larger bioaggregates (granules with diameter >=350 µm) and in smaller bioaggregates (diameter < 350 µm). The ratios calculated were based on the DNA-RPKM values of the *ppk1* genes in each clade, the total DNA-RPKM values of the *ppk1* genes in all the five clades was taken as 1, and the ratio of each specific clade was normalized accordingly.

文本

描述已自动生成

**Figure S2.** Maximum likelihood phylogenetic tree between the Competibacteraceae MAG GAO1 and 38 reference draft genomes in the family of Competibacteraceae. The tree was constructed using RAxML with 100 bootstraps based on a set of 120 concatenated universal single-copy proteins (Parks et al., 2018), and bootstrap values ≥0.9 are shown on the branches. The reference genome (represented by the accession number of GCA 00322995) of Thiohalomonadales was used as the outgroup. Reference genomes annotated at the genus level of *Competibacter* are labeled in orange, and reference genomes of the *Competibacter* lineage (Brand et al., 2019; Mcllroy et al., 2014; Mcllroy et al., 2015) beyond the genus of Competibacter are highlighted in blue. All references genomes annotatable at the genus level are italicized and corresponding genus names are shown in parentheses.

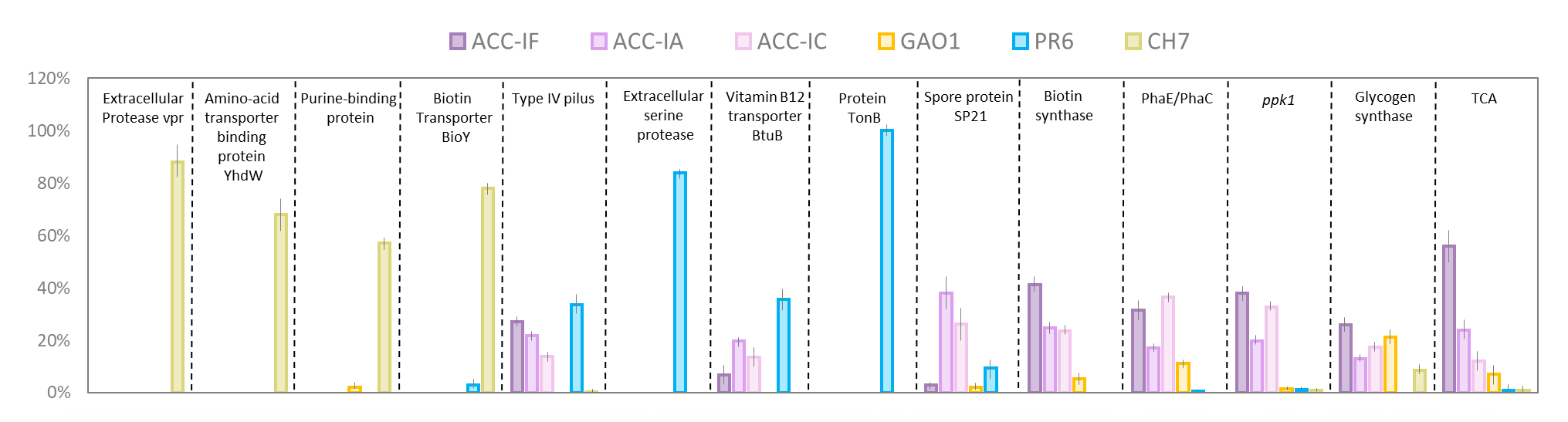
A screenshot of a cell phone

Description automatically generated

**Figure S3**. Denitrification gene expression under different redox conditions. The relative gene expression under anoxic and aerobic conditions was normalized to the RNA-RPKM value under anaerobic conditions. The relative gene expression is averaged across propionate and acetate fed cycles. Blue indicates up-regulation and red indicates down-regulation, relative to anaerobic conditions. Abundances of the gene transcripts expressed by each MAG are indicated by the size of the bubbles. 0- Nitrate/Nitrite transporter, 1- Nitrate reduction, 2- Nitrite reduction, 3- Nitric oxide reduction, and 4- Nitrous oxide reduction.



**Figure S4**. Gene expression patterns (RPKM and relative) of major carbon, electron transport, EPS production and P metabolic pathways by each dominant MAG. The 1st column on the left represents the RNA-RPKM values of the genes involved under the anaerobic period, and colors changing from purple to red represents the increase of RPKM values. The 2nd and 3rd columns represent relative gene expression under the anoxic and aerobic conditions, normalized by the RNA-RPKM values under the anaerobic condition. MAGs with missing genes in the pathways are marked as blank in the heatmap.



**Figure S5**. RNA-RPKM based quantification of gene transcripts associated with the key metabolism in each of the 6 identified highly active MAGs. Gene transcripts expressed by different MAGs were presented in different colors, and the height of each bar represents the proportion of gene transcripts expressed by each MAG. The RNA-RPKM values were averaged across the six metatranscriptome datasets. The specific RNA-RPKM values of genes associated with the metabolism in this figure can be found in the appendix file 3.



**Figure S6**. CAZy gene expression in the 6 dominant MAGs (in terms of gene expression). The 1st column on the left summarizes the RNA-RPKM values of each gene under the anaerobic period, where the size of each bubble represents the RNA-RPKM values. The 2nd and 3rd columns represent relative gene expression under the anoxic and aerobic conditions, normalized by the RNA-RPKM values under the anaerobic condition. GH: glycoside hydrolases, CBM: carbohydrate-binding module



**Figure S7**. Distribution of the aggregate size characterized via particle size analyzer. Blue indicate volume %, and green indicates cumulative volume %.

**Supplementary Information – References**

Brand, V. R., Crosby, L. D., & Criddle, C. S. (2019). Niche differentiation among three closely related Competibacteraceae clades at a full-scale activated sludge wastewater treatment plant and putative linkages to process performance. *Applied and environmental microbiology*, **85**(5).

McIlroy, S.J., Albertsen, M., Andresen, E.K., Saunders, A.M., Kristiansen, R., Stokholm-Bjerregaard, M.S., Neilsen K.L., and Nielsen P.H. (2014) 'Candidatus Competibacter'-lineage genomes retrieved from metagenomes reveal functional metabolic diversity. *ISME J* **8**: 613-624.

McIlroy, S.J., Nittami, T., Kanai, E., Fukuda, J., Saunders, A. M., & Nielsen, P. H. (2015). Re‐appraisal of the phylogeny and fluorescence in situ hybridization probes for the analysis of the Competibacteraceae in wastewater treatment systems. *Environmental Microbiology Reports* **7**:166-174.

Maldarelli, G. A., Piepenbrink, K. H., Scott, A. J., Freiberg, J. A., Song, Y., Achermann, Y., ... & von Rosenvinge, E. C. (2016). Type IV pili promote early biofilm formation by Clostridium difficile. *FEMS Pathogens and Disease*, **74:** ftw061.

Pohlschroder, M., & Esquivel, R. N. (2015). Archaeal type IV pili and their involvement in biofilm formation. *Frontiers in microbiology*, **6:** 190.

Shultis, D.D., Purdy, M.D., Banchs, C.N., Wiener, M.C. (2006) Outer Membrane Active Transport: Structure of the BtuB:TonB Complex. *Science* **312**: 1396-1398.