



Genome-Resolved Metagenomics of a Photosynthetic Bioreactor Performing Biological Nutrient Removal

Elizabeth A. McDaniel,^a Renske Wever,^{c,d} Ben O. Oyserman,^{c,d} Daniel R. Noguera,^{b,e} Katherine D. McMahon^{a,b}

^aDepartment of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA

^bDepartment of Civil and Environmental Engineering, University of Wisconsin—Madison, Madison, Wisconsin, USA

^cDepartment of Microbial Ecology, Netherlands Institute of Ecological Research, Wageningen, The Netherlands

^dBioinformatics Group, Wageningen University and Research, Wageningen, The Netherlands

^eDOE Great Lakes Bioenergy Research Center, Madison, Wisconsin, USA

ABSTRACT Enhanced biological phosphorus removal (EBPR) is an economically and environmentally significant wastewater treatment process for removing excess phosphorus by harnessing the metabolic physiologies of enriched microbial communities. We present a genome-resolved metagenomic data set consisting of 86 metagenome-assembled genome sequences from a photosynthetically operated lab-scale bioreactor simulating EBPR.

A photobioreactor performing light-supported enhanced biological phosphorus removal (EBPR) was seeded from two lab-scale wastewater treatment systems, a photosynthetic nitrifying enrichment culture and an activated sludge EBPR bioreactor designed to enrich for the polyphosphate-accumulating organism “*Candidatus Accumulibacter phosphatis*” (1, 2). The photobioreactor was designed to achieve phosphorus removal and denitrification without the addition of mechanical aeration by providing light (2). Three samples were collected during cycles 87, 103, and 129. DNA was extracted as described (2), and 100 ng of DNA was sheared to 300 bp using the Covaris LE220 ultrasonicator and size selected with solid-phase reversible immobilization (SPRI) beads (Beckman Coulter). The fragments were ligated with end repair, A-tailing, and Illumina-compatible adapters (IDT, Inc.) using the KAPA-Illumina library preparation kit (Kapa Biosystems). The libraries were quantified using the Kapa Biosystems next-generation sequencing library quantitative PCR (qPCR) kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were prepared using the v4 TruSeq paired-end cluster kit and the Illumina cBot instrument to create a clustered flow cell for sequencing. Shotgun metagenomic sequencing was performed on the Illumina HiSeq 2500 platform with the TruSeq sequencing-by-synthesis (SBS) sequencing kit, followed by 2 × 150-bp indexing. All metagenomic libraries consist of approximately 50 million 150-bp paired-end reads (table of metagenome information available at <https://doi.org/10.6084/m9.figshare.14164307>).

Default parameters were used for all software unless otherwise specified. The raw metagenomic reads were quality filtered and trimmed using bbduk as part of the BBTools v38.07 suite (3). All three metagenomic samples were coassembled and individually assembled into contigs using SPAdes v3.9.0 with the metagenomic option (4). Metagenomic reads from each sample were mapped against all assemblies using BBMap as part of the BBTools v38.07 suite with a 95% sequence identity cutoff (3). The assembled contigs from each assembly were binned into population genomes using MetaBAT2 v2.12.1 using only contigs larger than 1,000 bp (5). The bins were dereplicated using dRep v2.4.2 (6) to obtain the highest-quality representative set of genomes. This resulted in a total of 86 nonredundant species-resolved genome sequences, described in Table 1 and summarized in Fig. 1. All genome

Citation McDaniel EA, Wever R, Oyserman BO, Noguera DR, McMahon KD. 2021. Genome-resolved metagenomics of a photosynthetic bioreactor performing biological nutrient removal. *Microbiol Resour Announc* 10: e00244-21. <https://doi.org/10.1128/MRA.00244-21>.

Editor J. Cameron Thrash, University of Southern California

Copyright © 2021 McDaniel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Elizabeth A. McDaniel, elizabethmcd93@gmail.com.

Received 12 March 2021

Accepted 31 March 2021

Published 6 May 2021

TABLE 1 Genome accession numbers and statistics^a

Relative abundance on date (mo/day/year)																		
Isolate name	Assembly no.	Code	GTD& classification	Complete- contamination (%)	Genome size (mbp)	Avg contig size (bp)	% GC content	No. of rRNAs	No. of tRNAs	No. of 5S rRNAs	No. of 16S rRNAs	No. of 23S rRNAs	Avg abundance	BioSample accession no.	GenBank accession no.	BioProject accession no.		
UWPOB_ACID01	GCA_017304135.1	ACID01	d_Bacteriopl._Acidobacteriales; C_Acidobacteriieno_Bryohacteriales; f_Bryohacteriaceae;g_S;	98.4	2.9	6.49	365	17,791	64.5	3	48	1	1	0.04386	0.06137	0.15309	0.08611	
UWPOB_ACID02	GCA_017303951.1	ACID02	d_Bacteriopl._Acidobacteriales; C_Blastocatellia_o_UBA755ef_39_;	100	2.1	6.15	46	133,664	34.5	0	83	0	0	0.01576	0.0143	0.30104	0.11037	
UWPOB_ARMA1	GCA_017303935.1	ARMA1	d_Bacteriopl._Armatimonadetes; C_Fimbrimonadales; f_Fimbrimonadaceae_g_UBA2391;	100	0	2.56	5	511,480	53.2	3	47	1	1	1.87386	0.86294	0.31483	0.10821	
UWPOB_BACT1	GCA_017303995.1	BACT1	d_Bacteriopl._Bacteroidetes_o_f_;	100	0	4.31	137	31,460	41.5	2	40	2	0	0.12322	0.02324	0.02244	0.07763	
UWPOB_BACT2	GCA_017303975.1	BACT2	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_AKH767f_2-12-fULL-35-159_35_;	95.7	1	3.47	75	46,223	35.2	3	32	1	1	0.12227	0.02794	0.02315	0.05779	
UWPOB_BACT3	GCA_017303895.1	BACT3	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_AKH767f_2-12-fULL-35-159_35_;	99.1	0.7	3.61	35	103,001	37.3	3	33	1	1	0.10972	0.06722	0.02034	0.06576	
UWPOB_CHIT1	GCA_017303905.1	CHIT1	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_AKH767f_2-12-fULL-35-159_35_;	100	0	3.95	7	564,856	36.8	4	33	2	1	0.08039	0.12427	0.26855	0.15774	
UWPOB_CHIT2	GCA_017303835.1	CHIT2	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae_g_S;	99.5	0.3	3.26	2	1,631,117	42.5	3	36	1	1	2.73174	1.62266	3.05499	2.4698	
UWPOB_CHIT3	GCA_017303415.1	CHIT3	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae_g_S;	100	0	3.72	3	1,239,140	46	1	37	0	1	0	0.64663	0.61427	0.58909	0.61668
UWPOB_CHIT4	GCA_017303335.1	CHIT4	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae;	99	0.3	4.01	7	572,270	49.5	3	42	1	1	0.64336	0.73815	1.22698	0.87016	
UWPOB_CHIT5	GCA_017303455.1	CHIT5	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae_g_J0085_5_;	95.8	0	4.55	57	79,888	43	0	40	0	0	0.0768	0.06206	0.05275	0.06387	
UWPOB_CHIT6	GCA_017303855.1	CHIT6	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae_g_J0085_5_;	99.5	0.7	4.62	31	149,027	46.1	1	37	1	0	0	0.23357	0.2022	0.15554	0.1971
UWPOB_CHIT7	GCA_017303395.1	CHIT7	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Chitinophagae_g_Talbella_B;	100	6.3	3.76	45	83,566	41.9	0	33	0	0	0	0.12824	0.08724	0.02842	0.0813
UWPOB_CHIT8	GCA_017303475.1	CHIT8	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Saprospiraceae_g_S;	98.5	0.5	6.04	53	114,008	47.9	3	40	1	1	0.83323	1.32521	1.47064	1.21038	
UWPOB_CHIT9	GCA_017303875.1	CHIT9	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Chitinophagales; f_Saprospiraceae_g_J0085_5_;	99	0.5	5.78	64	90,342	50.4	0	43	0	0	0	0.12036	0.15272	0.42621	0.2331
UWPOB_CYTO1	GCA_017303805.1	CYTO1	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Cytophagales; f_Cytophagaceae_g_ELB16-189_;	98.5	0.3	4.29	13	329,632	45.1	1	35	1	0	0	0.03105	0.06252	0.27364	0.1224
UWPOB_CYTO2	GCA_017303795.1	CYTO2	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Cytophagales; f_Cytophagaceae_g_ELB16-189_;	100	0	4.87	27	180,295	42.2	2	34	1	0	1	0.06209	0.0432	0.20485	0.10338
UWPOB_FLAV01	GCA_017302935.1	FLAV01	d_Bacteriopl._Bacteroidetes; C_Bacteroidao_Flavobacteriales; f_Flavobacteriaceae;	93.2	0.8	3.09	264	11,718	34	2	40	2	0	0	0.04808	0.01753	0.05116	0.03892

(Continued on next page)

TABLE 1 (Continued)

(Continued on next page)

TABLE 1 (Continued)

(Continued on next page)

TABLE 1 (Continued)

relative abundance on date (mo/day/yr)	Relative abundance												BioProject accession no.						
	contaminant (%)			Genome size (mbp)			Avg contig size (bp)			% GC			No. of rRNAs	No. of 16S rRNAs	No. of 5S rRNAs	No. of 23S rRNAs	Avg abundance	bioSample accession no.	GenBank accession no.
JWPOB_RICK2	Assembly accession no.	Code	GTDB classification	Complete-ness (%)	Contamination (%)	Genome size (mbp)	contigs (bp)	contig size (bp)	content	% GC	No. of rRNAs	No. of 16S rRNAs	No. of 5S rRNAs	No. of 23S rRNAs	bioSample accession no.	GenBank accession no.	BioProject accession no.		
JWPOB_SPHING1	GCA_017302775.1	SPHING1	d_Bacteriopl. Proteobacteria; c_Alphaproteobacteria; o_Rickettsialesf_ Rickettsiaceae;	99.5	1	1.33	5	264,993	34.7	0	34	0	0	0	0.03107	0.0695	0.08658	0.1082	SAMN18059883 JAFLD0000000000 PRINA704939
JWPOB_SPHING2	GCA_017302615.1	SPHNG2	d_Bacteriopl. Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae; g_Novosphingiobiums—	98.5	5.1	3.16	556	5,684	61.6	0	48	0	0	0	0.06166	0.0599	0.01391	0.04516	SAMN18059885 JAFLD0000000000 PRINA704939
JWPOB_ALPHA2	GCA_017302575.1	ALPHA2	d_Bacteriopl. Proteobacteria; c_Alphaproteobacteria_UBA998; f_UBA30029_g_S—	98.9	0	1.86	1	1,861,815	50.8	3	39	1	1	1	0.47594	0.19689	0.04249	0.23844	SAMN18059886 JAFLD0000000000 PRINA704939
JWPOB_BURK1	GCA_017302655.1	BURK1	d_Bacteriopl. Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Burkholderiaeag_S—	100	0	4.44	203	21,857	67.3	2	48	1	1	0	0.37852	0.42244	0.55223	0.45106	SAMN18059887 JAFLD0000000000 PRINA704939
JWPOB_OTTO1	GCA_017302725.1	OTTO1	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderiales; f_Burkholderiaeag_Ottowia; s_Orcowia	95.8	0	2.87	325	8,835	68.6	3	52	1	1	1	0.76562	0.83948	0.30961	0.63823	SAMN18059888 JAFLD0000000000 PRINA704939
JWPOB_RUBRI1	GCA_017302495.1	RUBRI1	d_Bacteriopl. Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Burkholderiaeag_Rubrivivax—	81	0.8	4.51	553	8,152	71.1	3	56	1	1	1	1,62184	1,49067	0.45613	1.18955	SAMN18059889 JAFLD0000000000 PRINA704939
JWPOB_RUBRI2	GCA_017302505.1	RUBRI2	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderiales; f_Burkholderiaeag_Rubrivivax—	100	0	5.5	555	9,913	68.4	4	52	2	1	1	0.21114	0.24105	0.10683	0.18634	SAMN18059890 JAFLD0000000000 PRINA704939
JWPOB_VTR1	GCA_017302465.1	VTR1	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderiales; f_Burkholderiaeag_Vitreoscilla_A—	95.8	0	3.68	513	7,177	68.4	1	42	1	0	0	0.07393	0.05699	0.17722	0.10268	SAMN18059891 JAFLD0000000000 PRINA704939
JWPOB_BURK2	GCA_017302475.1	BURK2	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderialesf_Palsia-1005g—	79.2	0	2.71	498	5,448	69.6	0	34	0	0	0	0.10158	0.12158	0.111145	0.11154	SAMN18059892 JAFLD0000000000 PRINA704939
JW10-POB	GCA_017302555.1	CAPIB	d_Bacteriopl. Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocyclaceae;	100	0	4.4	88	49,949	62.5	0	41	0	0	0	0.52449	0.60697	0.65232	0.59459	SAMN18059893 JAFLD0000000000 PRINA704939
JW12-POB	GCA_017302435.1	CAPID	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderiales; f_Rhodocyclaceae;	98.1	0.7	4.56	91	50,155	62.7	0	46	0	0	0	0.35414	0.32156	0.05167	0.24246	SAMN18059894 JAFLD0000000000 PRINA704939
JW13-POB	GCA_017302415.1	CAPIF	d_Bacteriopl. Proteobacteria; c_Gammoproteobacteria; o_Burkholderiales; f_Rhodocyclaceae;	97.1	4.4	545	245	22,232	65.9	1	48	0	1	0	1,65004	0.70231	0.09466	0.81567	SAMN18059895 JAFLD0000000000 PRINA704939

(Continued on next page)

TABLE 1 (Continued)

Isolate name	Assembly accession no.	Code	GTDB classification	Relative abundance on date (mo/day/yr)															
				Complete- ness (%)	Contami- nation (%)	Genome size (mbp)	contigs size (bp)	% content	% GC	No. of rRNAs	No. of tRNAs	No. of rRNAs	No. of tRNAs	Avg abundance	No. of BioSample accession no.	BioProject accession no.	GenBank accession no.		
UW11_POB	GCA_017302385.1	CAPIC	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;	100	0	4.63	144	32,127	61.2	0	47	0	0	12,2117	11,8669	10,2307	SAMN118059898 JAFLDR000000000 PRINA704939		
UW8_POB	GCA_017302345.1	CAPIA	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;	100	0	4.26	49	86,993	64.2	0	51	0	0	6,34328	6,19256	0,4336	4,32315	SAMN118059896 JAFLDN000000000 PRINA704939	
UW9_POB	GCA_017302455.1	CAPIA	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;	100	0	4.93	126	39,138	64.1	1	46	1	0	0,55782	0,54654	0,15933	0,42123	SAMN118059897 JAFLDO000000000 PRINA704939	
UWPOB_DECHI	GCA_017302355.1	DECHI	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;	91.5	8.1	3.69	279	13,238	65	2	40	0	1	2,09698	3,33708	3,17482	2,86963	SAMN118059899 JAFLDQ000000000 PRINA704939	
UWPOB_SULF1	GCA_017302275.1	SULF1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;g_Sulfurales;—	98.3	1.1	4.41	88	50,106	64	3	47	2	1	0	0,76164	0,92706	2,2433	1,31067	SAMN118059900 JAFLDR000000000 PRINA704939
UWPOB_THAU1	GCA_017302295.1	THAU1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;g_Thaerae;—	98.7	1.4	4.11	56	73,417	66.1	3	52	3	0	0	0,5635	0,57917	0,40828	0,51699	SAMN118059901 JAFLDS000000000 PRINA704939
UWPOB_ZOG1	GCA_017309145.1	ZOG1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales; f_Rhodocylaceae;g_Zoogleas;—	83.3	4.2	4.91	428	11,461	65.7	0	58	0	0	0,06958	0,06569	0,07688	0,07071	SAMN118059903 JAFLD000000000 PRINA704939	
UWPOB_ZOG2	GCA_017302315.1	ZOG2	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Burkholderiales;	95.2	1.8	4.52	165	27,407	64.5	0	55	0	0	0,12938	0,14704	0,21305	0,16316	SAMN118059902 JAFLDT000000000 PRINA704939	
UWPOB_PARA1	GCA_017302255.1	PARA1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Enterobacterales; f_Alteromonadaceae;	100	4.2	4.03	77	52,358	46.2	0	50	0	0	0,20334	0,02657	0,00492	0,07828	SAMN118059904 JAFLDV000000000 PRINA704939	
UWPOB_GAMMA1	GCA_017302205.1	GAMMA1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteriia_GCA_2729095f_GCA_2729095g_5—	79.2	1	2.51	456	5,508	69.2	1	34	1	0	0,05103	0,07089	0,10094	0,07429	SAMN118059905 JAFLDW000000000 PRINA704939	
UWPOB_RHODAN1	GCA_017302215.1	RHODAN1	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Rhodanobacteraceae;	91.7	0	3.5	323	10,843	68.8	0	49	0	0	0,144	0,20595	0,16775	0,17257	SAMN118059906 JAFLDX000000000 PRINA704939	
UWPOB_AQU11	GCA_017302155.1	AQU11	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae;	98.3	1.6	3.98	113	35,212	61.5	0	46	0	0	0,28393	0,24393	0,06593	0,19793	SAMN118059907 JAFLDY000000000 PRINA704939	
UWPOB_LYS01	GCA_017302135.1	LYS01	d_Bacteriap__Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae;	99.6	1	4.34	35	123,959	64.1	2	48	1	1	0,12091	1,20058	0,96205	1,12451	SAMN118059908 JAFLDZ000000000 PRINA704939	

(Continued on next page)

TABLE 1 (Continued)

Isolate name	Assembly accession no.	Code	GTDB classification	Relative abundance on date (mo/day/yr)																		
				Completeness (%)	Contamination (%)	Genome size (mbp)	No. of contigs	% GC content (bp)	No. of rRNAs	No. of tRNAs	No. of sRNAs	Avg abundance (7/16/2015)	Avg abundance (7/24/2015)	Avg abundance (8/6/2015)	Avg abundance (7/16/2015)	Avg abundance (7/24/2015)	Avg abundance (8/6/2015)	No. of GenBank accession no.	BioProject accession no.	BioSample accession no.	GenBank accession no.	BioProject accession no.
UWPOB_PSEUDO1	GCA_017302035.1	PSEUDO1	d_Bacteriapl_Proteobacteria; c_Gammaproteobacteria; f_Xanthomonadales; f_Xanthomonadaceae;	95.8	0	2.91	125	23,300	69.3	1	48	1	0	0	1,50336	2,17593	0.47299	1.38409	SAMN18059909	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_THERMO1	GCA_017302095.1	THERMO1	g_Pseudoxanthomonas; d_Bacteriopl_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae; g_Xanthomonadaceae; 9_Xanthomonadales;	76.4	2.1	2.31	299	7,734	69.7	0	39	0	0	0	0.21965	0.25626	0.05906	0.17832	SAMN18059911	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_THERMO2	GCA_017302075.1	THERMO2	d_Bacteriopl_Proteobacteria; c_Gammaproteobacteria; o_Xanthomonadales; f_Xanthomonadaceae;	87.5	0.5	2.77	245	11,303	69.9	0	41	0	0	0	0.79368	1,14049	0.4586	0.79759	SAMN18059910	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_SPIRO1	GCA_017302045.1	SPIRO1	d_Bacteriopl_Spirochaeta; c_Lepotospirae_Turbellariales; f_Turnerellaeag_Turnerellales; g_Thermomonas; 9_Thermomonas;	86.5	0	3.78	412	9,184	53.8	2	31	2	0	0	0.03758	0.03209	0.0621	0.04392	SAMN18059912	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_SPIRO2	GCA_017302015.1	SPIRO2	d_Bacteriopl_Spirochaeta; c_UBA1235o_f_39_5_	87.5	0	5.43	173	31,368	65.2	2	41	0	1	1	0.09083	0.09013	0.07872	0.08656	SAMN18059913	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_VERRUCO1	GCA_017302195.1	VERRUCO1	d_Bacteriopl_Verrucomicrobiota; c_Verrucomicrobia; o_Pedosphaeraesf_UBA9464; 9_UBA9464s_	81	8.3	4.86	848	5,726	67.1	0	35	0	0	0	0.04712	0.07195	0.01614	0.04507	SAMN18059915	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_VERRUCO2	GCA_01730115.1	VERRUCO2	d_Bacteriopl_Verrucomicrobiota; c_Verrucomicrobia; o_Pedosphaeraesf_UBA9464; 9_UBA9464s_	98.6	3.5	753	112	67,228	61.5	3	50	1	1	1	0.14886	0.18964	0.1474	0.16196	SAMN18059914	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_VERRUCO3	GCA_017301995.1	VERRUCO3	d_Bacteriopl_Verrucomicrobiota; c_Verrucomicrobia; o_Verrucomicrobiales; f_Akkermansiaceag_UBA1315s_	100	0	7.04	68	103,585	62.3	2	60	2	0	0	0.30515	0.26629	0.07792	0.21646	SAMN18059916	JAFLE000000000	JAFLE000000000	PRINA704939
UWPOB_VERRUCO4	GCA_017301975.1	VERRUCO4	d_Bacteriopl_Verrucomicrobiota; c_Verrucomicrobia; o_Verrucomicrobiales; f_Verrucomicrobiaceae;	97	0	6.9	97	71,172	60.3	0	54	0	0	0	0.17845	0.15859	0.07204	0.13636	SAMN18059917	JAFLE000000000	JAFLE000000000	PRINA704939

^a Genome information for 86 metagenome-assembled genomes is shown. The column labeled "Isolate name" lists the name submitted to NCBI, and that labeled "Code" lists the codes used in Fig. 1. Classifications were assigned with GTDB-tk and "Candidatus Accumulibacter phosphatis" clade designations based on comparisons to ptk7 clones and publicly available reference genome sequences. "Candidatus Accumulibacter phosphatis" genomes are named alpha numerically according to previous nomenclature; all other genomes are named UWPOB_CODE. The genome statistics for completeness, contamination, GC content, genome size, and number of contigs were performed with CheckM v1.1.2. The rRNA and tRNA genes were predicted with Barnap v0.9, as part of the Prokka v1.13.7 package. The relative abundance was calculated using coverM v0.4.0 using the relative_abundance method for the three metagenomic samples, taken on 16 July 2015, 24 July 2015, and 6 August 2015 toward the end of the enrichment period, and averaged together across all three samples for an average relative abundance calculation.

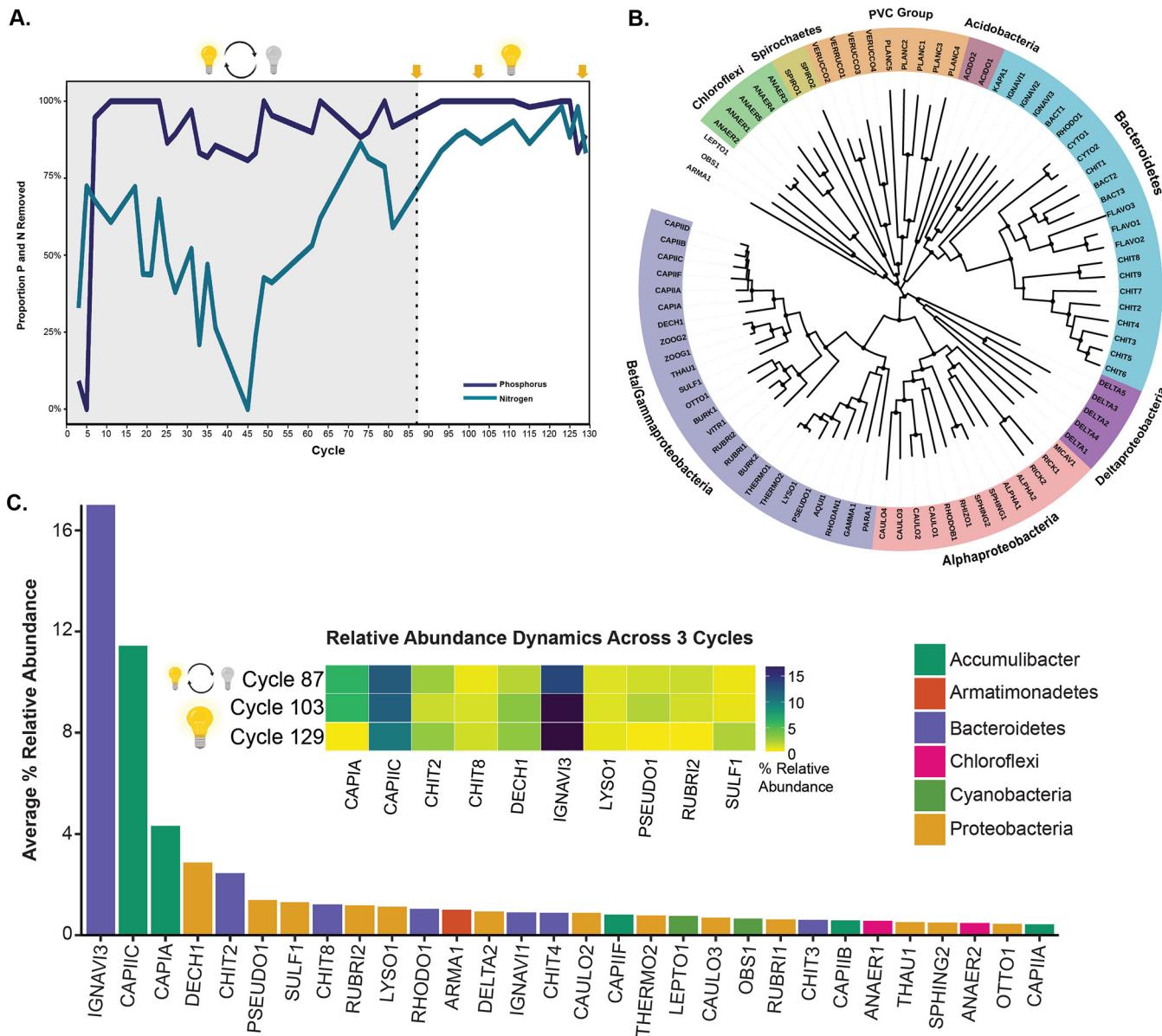


FIG 1 Overview of reactor performance and assembled MAGs. (A) Proportions of nitrogen (blue) and phosphorus (purple) removed during reactor operation. Cycle length refers to the hydraulic residence time of 0.625 days (2). The reactor was operated in two phases—in the first phase (gray), light was cycled on and off, and in the second phase (white), light was provided continuously. The arrows denote the time points at which metagenomic sequencing was performed. (B) Phylogenetic tree of all 86 assembled MAGs constructed from a concatenated alignment of 120 single-copy marker genes from GTDB-tk v0.3.2 (9) that were identified and aligned using HMMER v3.2.1 (11). The tree was constructed using RAxML v8.2.12 with 100 rapid bootstraps (12) and viewed using iTOL v5.7 (13). (C) Relative abundance and population dynamics of abundant species represented by assembled MAGs. The bar graph represents the average relative abundance of the top 30 most abundant species across all three samples. Relative abundance calculations were made using coverM v0.4.0 with the relative_abundance method and averaged across all three samples. The inset heat map shows the population dynamics of the top 10 most abundant species, representing the relative abundance of that MAG in that sample.

sequences were checked for uniform differential coverage using uBin v0.9.14 (7). All genome statistics were calculated with CheckM v1.1.2 (8). Taxonomical classifications were made based on the Genome Taxonomy Database (GTDB) using GTDB-Tk v0.3.2 (9, 10). Relative abundance calculations were performed with coverM v0.4.0 using the relative_abundance calculation method. The methods and phylogenetic tree for assigning clades of six “*Ca. Accumulibacter phosphatis*” draft genome sequences based on the *ppk1* locus are available at <https://doi.org/10.6084/m9.figshare.14164478>.

Data availability. The raw metagenomes for the 3 samples and genome assemblies for all 86 metagenome-assembled genomes (MAGs) are available at NCBI GenBank

under BioProject accession number [PRJNA704939](#). The metagenomes are available under the SRA accession numbers [SRR13786854](#), [SRR13786855](#), and [SRR13786856](#).

ACKNOWLEDGMENTS

This work was supported by funding from the National Science Foundation (MCB-1518130) to K.D.M. and D.R.N.

We thank the University of Wisconsin—Madison Biotechnology Center DNA Sequencing Facility for Illumina sequencing services. This research was performed in part using the Wisconsin Energy Institute computing cluster, which is supported by the Great Lakes Bioenergy Research Center as part of the U.S. Department of Energy Office of Science (DE-SC0018409).

REFERENCES

- Martín HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyripides NC, Blackall LL, McMahon KD, Hugenholtz P. 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* 24:1263–1269. <https://doi.org/10.1038/nbt1247>.
- Oyserman BO, Martirano JM, Wipperfurth S, Owen BR, Noguera DR, McMahon KD. 2017. Community assembly and ecology of activated sludge under photosynthetic feast-famine conditions. *Environ Sci Technol* 51:3165–3175. <https://doi.org/10.1021/acs.est.6b03976>.
- Bushnell B, Rood J, Singer E. 2017. BBMerge—accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056. <https://doi.org/10.1371/journal.pone.0185056>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotnik AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>.
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
- Bornemann TLV, Esser SP, Stach TL, Burg T, Probst AJ. 2020. uBin—a manual refining tool for metagenomic bins designed for educational purposes. *bioRxiv*. <https://doi.org/10.1101/2020.07.15.204776>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7: e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
- Stamatakis A. 2014. The RAxML v8.0.X manual. <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>.
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.