

Toward energy self-sufficient wastewater treatment: insights into biological nutrient removal using low dissolved oxygen

By

Natalie Keene Beach

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Civil and Environmental Engineering)

At the University of Wisconsin-Madison

2019

Date of final oral examination: July 1, 2019

Committee Members:

Daniel R. Noguera, Professor, Civil and Environmental Engineering

Katherine D. McMahon, Professor, Civil and Environmental Engineering

Greg Harrington, Professor, Civil and Environmental Engineering

Karthik Anantharaman, Professor, Bacteriology

Garret Suen, Professor, Bacteriology

This page intentionally left blank

Abstract

Many water resource recovery facilities (WRRF) around the world utilize biological processes for effective removal of nitrogen (N) from municipal waste; since biologically available N has been recognized as a significant pollutant in wastewater, is toxic to aquatic life and contributes to eutrophication. Over the past 30 years, discoveries of novel N-cycling microorganisms and biological pathways have been especially influential to energy-conscious WRRF research. Today, most biological N removal processes rely on extensive aeration for nitrification and organic carbon for denitrification; however, the high cost and loss of energy generating carbon associated with this type of treatment has propelled engineers and scientists to develop novel biological N treatment strategies that employ many of the newly discovered microorganisms and pathways. While energy-saving strategies are promising and have been employed full-scale, the microorganisms involved in some of these micro-aerobic N-cycling engineered ecosystems remain elusive.

We have shown at the pilot-scale (treating full-scale primary effluent) that we can slowly transform an existing high-oxygen N-cycling microbial community to low oxygen concentrations, without loss of effluent quality, and with improved N removal efficiency [9]. Separate work with a bench-scale Madison Micro-Aerobic/AnoXic (MAD MA-AX) bioreactor has also shown that we can achieve up to 90% N removal with minimal aeration to treat the reject water from a struvite harvesting process. In both cases, we describe the key microorganisms contributing to N removal and discuss the factors that may lead to adaptation to microaerobic conditions.

To my incredible husband, John, and parents, Jack and Nancy.

*“It always seems impossible until it’s done.”
—Nelson Mandela*

Acknowledgements

Wow! I cannot believe I have made it to this moment. These past 6 years have been quite a marathon, and I am finally passing the finish line with incredible strength, a passion for science, and a deeper understanding of myself. I am so grateful for everything this journey has had to offer, and I wish I could thank everyone that made an impact on my life during all these years. Here, I would like to take the time to highlight some very special people who have made this all possible.

First, I would like to express my deepest gratitude to my advisor, Dr. Daniel Noguera. Thank you for taking a chance on me all those years ago by giving me a call to offer me the research assistantship in your lab. I will never forget your incredible mentorship, encouragement, guidance, and patience. You have both inspired me and pushed me to think outside of my comfort zone, and for that I have grown beyond my wildest imagination.

I would also like to thank my committee members (Katherine “Trina” McMahon, Greg Harrington, Karthik Anantharaman, and Garret Suen) and Madison Metropolitan Sewerage District staff (Paul Nehm, Steve Reusser, Alan Grooms, and Matt Seib) for their thoughtful input, guidance and support over the past several years.

I have learned a tremendous amount and enjoyed working alongside members of both the Noguera and McMahon Research groups. I am so thankful to have had the opportunity to work with an extremely talented group of people, including (but not limited to) Dr. Pamela Camejo, Dr. Francisco Moya, Dr. Matthew Scarborough, Christopher Lawson (almost Dr.), Dr. Zach Oshlag, Dr. Ben Oyserman, Dr. Weiping

Zhang, Duygu Celik, Natalie Cook, Brian Owen, Rachel Stewart and Abel Ingle. I am so happy to have gained so many lifelong friends here. Dr. Matthew Scarborough: Thank you for being you and for always having my back. Your company has always brought me joy. I am so grateful to have had the opportunity to collaborate with a such an intelligent and thoughtful friend. I would also like to thank Dr. Francisco Moya and Dr. Pamela Camejo, for their positivity, enduring support, and friendship. Your zest for life is everything, and I can never get enough of being around you two. I cannot wait to visit you in Chile! Thank you also to Natalie Cook. It has been fun to share both a name and a birthday with you. You have always inspired me through your fearless and confident spirit. You taught me how to be better at conference networking, and I'm pretty sure I have a job now because of that! Cheers to many future WEFTEC reunions.

I would also like to thank the staff in the Department of Civil and Environmental Engineering. I am so happy to have had the opportunity to work with our lab manager, Jackie Cooper, over the last 6 years. I am so grateful for her support, patience, helping with learning new instruments, friendship and fun conversations. She not only taught me so much, but also made it so fun to be in the lab. I am also tremendously thankful for the hard work of numerous undergraduates who contributed to the research in this dissertation. It was a pleasure to work with Bradley Gliszinski, Sam Acker, Diana Barrera, Matt Dysthe, Alexandria Sabarots, Elizabeth Erb, Kate Dapolito, Cassie Ganos, Rahim Ansari, Jonnathan Garcia-Huerta, Tara Hawes, and Brandon Nutley.

To all my friends both back home in Colorado [and in Wisconsin]—thank you for being so kind and understanding when I needed to miss important events. These sacrifices were never easy, but you all were so gracious to me during this time.

I would especially like to thank Matt and Lindsay Garcia. You two are a light in my life. I could not imagine this experience without you. I will never forget our regular game nights, double dates, coffee dates, and holiday celebrations. Thank you for giving me so much love, so many laughs, and balance during my crazy time in graduate school.

I have never felt anything but unconditional support from my parents, Jack and Nancy Keene. Thank you for teaching me to be tenacious, push myself, and set high standards. You have fostered my competitive and curious spirit since the day I showed up, and I would not be the person that I am today without you. For that, your patience and constant encouragement, I am eternally grateful. I am also so very thankful for my wonderful, newly acquired family, Jim and Peggy Heimermann; Heather, Tim, Gabe, and Alea Edwards; Tom, Sam, and Joey Beach; LeeAnn, Marshal, and Blake Schmidt; Rebekah Beach and Daniel Beach. Thank you so much for welcoming me into your family with open arms. I am so lucky to have you all in my life!

Finally, and most importantly, I would like to thank my husband and best friend, John Beach. Thank you for believing in me, for picking me up whenever I was down, for pushing me when I needed it most, and for enduring the unconventional 24/7 life that research and graduate school can demand. Your unwavering support, encouragement, humor, and brilliant cooking skills have made graduate school some of the best years of my life. Thank you for now trekking across the country with me as I begin my career post-PhD. You are my rock, and without you, this dissertation would not have been possible.

This page intentionally left blank

Table of Contents

Abstract	ii
Acknowledgements.....	iv
List of Figures	xi
List of Supplementary Figures	xvi
List of Tables	xx
List of Supplementary Tables	xxi
Chapter 1.....	1
Global significance of wastewater treatment.....	2
Nitrogen removal	4
Nitrification.....	6
Denitrification	8
Anammox.....	8
Biological phosphorus removal.....	9
Energy requirement for biological nutrient removal	11
References.....	13
Chapter 2.....	24
Abstract	25
Introduction	26
Materials and Methods.....	28
Continuous flow pilot-scale reactor description and operation	28
Sample preparation and analysis.....	30
Kinetic tests	31
Floc and microcolony size measurements	31
16S ribosomal RNA gene tag sequencing.....	32
Energy Savings Calculations.....	34
Results	34
Reactor performance during step reductions in aeration	34
Extent of simultaneous nitrification and denitrification (SND) during reactor operation	37
Microbial adaptation to low-DO	39
Microbial community composition and dynamics	45
Energy Savings.....	54
Discussion	56
Conclusions	61

Acknowledgements.....	62
References	63
Supplementary Tables	71
Supplementary Figures	74
Supplementary Methods.....	78
Floc and microcolony sizes	78
Calculations of oxygen requirement and energy savings	79
Supplementary References	85
Chapter 3.....	86
Abstract	87
Introduction	88
Methods	89
Sample collection, processing, and DNA extraction	89
Design of primers to detect comammox ammonia monooxygenase gene amoA	91
Quantitative real-time polymerase chain reaction (qPCR)	91
Standard preparation for qPCR	95
Gradient PCR	96
Phylogenetic tree construction.....	98
Average nucleotide identity calculation.....	98
Results	99
Total comammox and clade-level comammox detection.....	99
Design of species-specific primer sets	105
Environmental detection of candidate comammox species	112
Discussion	115
Conflict of Interest	119
Author Contributions	119
Funding.....	119
Acknowledgments	119
References.....	120
Supplementary Tables	125
Supplementary Figures.....	133
Chapter 4.....	153
Abstract	154
Introduction	156
Results and Discussion	157

Characterization of reactor operation	157
Metagenomic sequencing, binning and quality	164
Phylogenetic analysis / Phylogenetic placement of MAGs.....	166
Microbial community abundance and gene expression.....	169
The four most transcriptionally active MAGs	170
The CHB metagenome	177
The AOB metagenome	194
The NOB metagenome	203
The AMX metagenome	211
Methods	220
Sequencing batch reactor description and operation	220
16S ribosomal RNA gene tag sequencing.....	222
Metagenomic DNA and RNA sequencing.....	223
Metagenomic data preparation, quality filtering, read recruitment, taxonomy, and phylogeny .	226
Genome annotation and homology search	228
Pangenome analysis / comparative genomics	229
Metatranscriptomic data preparation, quality filtering, read recruitment, and analysis	230
Gene expression time series data clustering.....	231
FNR sequence motif identification	232
Data availability.....	232
Conflict of Interest	233
Author Contributions	233
Funding.....	233
References.....	234
Supplementary Tables	245
Supplementary Figures.....	253
Chapter 5.....	263
Conclusions and Future Directions.....	264
References.....	268

List of Figures

- Figure 1-1.** Fifteen transformations among eight key nitrogen redox states (adapted from Kuypers et al. [14]). Red arrows represent an oxidation reaction that requires oxygen or another electron acceptor; blue arrows represent a reduction reaction, where organic carbon or other available electron donors are needed to complete the pathway. While nitrogen-transforming processes (upper-right) describe specific redox pathways; the microorganisms belonging to key functional groups are actually metabolically versatile and may contain multiple capabilities and/or may perform pathways simultaneously. 5
- Figure 1-2.** Evolving research of low oxygen biological nitrogen removal shows successful performance without a consensus identity to the key microbial players. [70, 73-79] 11
- Figure 2-1.** Continuous flow pilot-scale treatment train operated as a modified UCT process without nitrate recycling, simulating the configuration of the full-scale Nine Springs WWTP. The anaerobic (3 tanks), anoxic (1 tank), and aerobic (3 tanks) portions of the system were equipped with mechanical mixers. The internal recycle (QR1) returned biomass to the beginning of the system from the anoxic tank. Return activated sludge (QR2) was recycled from the bottom of the clarifier to the anoxic zone. Wasting (Qw) occurred from the last aerobic tank. 28
- Figure 2-2.** Pilot plant performance during the three operational phases. (A) Dissolved oxygen (DO), represented as a 20-point moving average time series from the three aerobic tanks. (B) Effluent Total Ammoniacal Nitrogen (TAN) and percent of simultaneous nitrification and denitrification (SND). (C) Effluent phosphate (P). (D) Activated sludge temperature. 35
- Figure 2-3.** Batch experiments of anoxic phosphorus uptake in the presence of nitrate or nitrite, (A) during high-DO operation, and (B) after stable low-DO operation. 38
- Figure 2-4.** (A) Comparison of specific rate of ammonia oxidation of activated sludge from the full-scale plant, and from the pilot-scale plant, after stable low-DO operation. Points represent experimental measurements of specific rates at different DO concentrations. Lines represent the best-fit of a Monod kinetic model to the experimental data. (B) Specific rates of phosphorus uptake of activated sludge from the full-scale plant and from the pilot-scale plant after stable low-DO operation (phase 3). Symbols represent experimental measurements of specific rates at different DO concentrations. A Monod kinetic model (solid line) could only be fitted to the data from the pilot-scale plant. A dashed line represents the average maximum specific rate for the full-scale plant. 41
- Figure 2-5.** Box plots comparing (A) floc size distribution (by median), and (B) AOB microcolony size distribution (by median) for the full-scale plant operating with high DO concentrations and the pilot plant during low-DO operation. The upper and lower bounds of the boxes represent the 25th and 75th percentiles, the whiskers denote the 10th and 90th percentiles, the circular symbols are the outliers, and the plus symbol represents the mean. 44
- Figure 2-6.** Box plots of the 30 most abundant core taxa classified at the genus level (by median) for (A) Pilot plant, and (B) Full-scale plant during the low-DO operational phase (phase 3). The names in the horizontal axis are the lowest assigned taxonomic rank. The upper and lower bounds of the boxes represent the 25th and 75th percentiles, the whiskers denote the maximum and minimum values, and the plus symbol represents the mean. 46
- Figure 2-7.** Balloon plots representing temporal variation in relative abundance of genera known to perform phosphorus cycling in EBPR reactors, (A) *Candidatus Accumulibacter* and (B) *Tetrasphaera*. For each genus, the pilot plant is described in the upper panel and full-scale plant in the lower panel. Relative abundance was determined based on the entire community. The shaded areas in

the upper panels denote the three operational phases, whereas the shaded areas in the lower panels indicate the meteorological seasons between days 120 and 490 of pilot plant operation. 48

Figure 2-8. Principle coordinate analysis (PCoA) ordination of next-generation sequence data comparing sample composition between full-scale and pilot-scale plants using Bray-Curtis dissimilarity matrices. Full-scale samples and pilot plant samples are represented as circles and triangles, respectively. The sample label represents the day of pilot plant operation that the sample was collected. The color gradient denotes the average DO for the pilot plant samples during each operational phase. (A) PCoA with all OTUs from the rarefied and filtered dataset considered. (B) PCoA with OTUs classified as the genera *Candidatus Accumulibacter*, and with relative abundance calculated using only the total number of reads in the *Ca. Accumulibacter* specific OTUs. (C) PCoA with relative abundance calculated using only the *Nitrosomonas* specific OTUs. (D) PCoA with relative abundance calculated using only the *Nitrospira* specific OTUs. 50

Figure 2-9. Balloon plots representing temporal variation in relative abundance of bacterial genera that participate in nitrification, (A) *Nitrosomonas*, (B) *Nitrospira*, (C) *Candidatus Nitrotoga*, and (D) *Nitrobacter*. For each genus, the pilot plant is described in the upper panel and full-scale plant in the lower panel. Relative abundance was determined based on the entire community. The shaded areas in the upper panels denote the three operational phases whereas the shaded areas in the lower panels indicate the meteorological seasons between days 120 and 490 of pilot plant operation. 52

Figure 2-10. Neighbor-joining consensus tree generated from an alignment of published 16S rRNA sequences and sequences retrieved in this study, rooting with *Nitrososvibrio*. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. The scale bar indicates a 2% sequence difference. Accession numbers are presented after the sequence names. 53

Figure 2-11. Estimated energy use reductions at the Nine Springs WWTP for 8 different DO scenarios.56

Figure 3-1. Visual representation of primer amplification regions with A) *Ca. N. nitrosa amoA*, B) *Ca. N. inopinata amoA*, and C) *Ca. N. nitrificans amoA*. The new primers, labeled with red arrows, were designed using full-length *amoA* sequences and novel primer binding region was discovered near the end of each *amoA* gene (approximately between 400 and 800 bp). Base pair locations are indicated in the primer names. Most publicly available published *amoA* sequences are partial length and are located between approximately 244 bp and 659 bp. Amplification with clade A-specific equimolar primer mixtures occur in this region, while amplification with the total comammox primers occur between 162 bp and 359 bp. 100

Figure 3-2. Abundance of *amoA* genes (gene copies/ng DNA) on a log-scale from comammox *Nitrospira* using the (A) total comammox primers, (B) clade A primers (C) clade B primers, (D) the newly designed *Ca. N. nitrosa* specific primers (E) *Ca. N. inopinata* specific primers and (F) *Ca. N. nitrificans* specific primers from various time series samples originating from five bioreactors. LS, lab-scale sequencing batch reactor; P_SBR, pilot-scale sequencing batch reactor; P_CF, continuous flow pilot plant; TRA, Trinity River WWTP; NS, Nine Springs WWTP. Error bars show the standard deviation of the triplicate samples. The color of each column indicates whether there was a true positive detection, an overestimated detection, or a false positive detection. The median limit of detection and quantification for each assay are represented by horizontal red and blue dotted lines, respectively. The shaded regions surrounding these limits represent the upper and lower bounds of the 95% confidence interval. 101

Figure 3-3. Neighbor-joining consensus tree generated from an alignment of full-length and near full-length *amoA* and *pmoA* gene sequences (> 600bp), rooting with archaea *amoA*. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. Bold sequence names were included in the new primer design. Blue, red, and green text are amplified by the *Ca.*

N. nitrificans, *Ca. N. nitrosa*, and *Ca. N. inopinata* primer sets, respectively. Asterisks indicate that the gene is amplified but with a reduced efficiency due to base pair mismatches. The scale bar indicates the number of nucleotide substitutions per site. Accession numbers are presented after the sequence names. Acronyms were used for groundwater well (GWW), wastewater treatment plant (WWTP), and membrane bioreactor (MBR). 108

Figure 3-4. Primer-target mismatch analysis for the newly designed primers and near full-length comammox clade A *amoA* greater than 600 bp. Predicted PCR efficiency is reported separately for the forward and reverse primers in addition to a combined amplification efficiency. Dots indicate a base match and letters indicate a mismatch, with the letter indicating which base is actually present in the sequence. 110

Figure 3-5. Matrix demonstrating the percent alignment of *amoA* genes (white to yellow to red gradient located on the bottom left portion of the matrix) and average nucleotide identity (ANI) of whole genome open reading frame alignments (white to pink gradient located on the top right portion of the matrix). Here, the percent genome ANI is shown in bold on the top of each square and fraction aligned is shown italicized and in brackets beneath the ANI value. 111

Figure 4-1. Bioreactor performance shown through: (A) A typical chemical profile during a single 8-hour cycle of operation; the plotted COD represents is the soluble fraction; and (B) a closeup of a 30 min time interval, defined by one aerobic and one anoxic cycle. RNA sampling were collected during time interval A and B, between 3-3.5 h and 4.5-5 h, respectively. In each interval, samples were collected during the aerobic period at 3 min and 7 min into the interval, indicated by '1' and '2' shaded regions. Additional samples were collected during the anoxic period at 13 min and 27 min, indicated by '3' and '4' shaded regions. 159

Figure 4-2. An 8-hour reactor cycle. (1) The system is slowly fed with the full-scale struvite harvester effluent during the first four hours of the cycle. During this time, conditions continuously cycle between a 10-minute micro-aerobic (< 0.2 mg O₂/L) and 20-minute anoxic period. (2) These intervals repeat for the next 3 ½ hours (between hours 4 - 7.5). (3) Then, the reactor is allowed to settle for 3 min, decants for 12 min and is idle for 15 additional min before the next cycle begins. 160

Figure 4-3. Phylogenetic tree generated from genome alignment of 16 most abundant MAGs recovered from the MAD MA-AX bioreactor. The tree includes MAGs recovered from this study (bold) and closely related genomes downloaded from the NCBI genome repository. The four most active MAGs (based on metatranscriptomics) are shown in bold orange text. Accession numbers for each genome are provided in parentheses. Bootstrap values are shown at the nodes where the value was greater than 50. The tree was constructed using RAxML based on a set of 37 concatenated universal single-copy marker genes. 167

Figure 4-4. Relative abundance and gene expression of 16 most abundant metagenomic assembled genomes (MAGs) recovered from the MAD MA-AX bioreactor. Relative abundance was determined by the number of DNA reads mapped to each MAG, normalized by the corresponding genome size. Relative expression (transcript abundance) was determined by the number of cDNA reads that mapped to each MAG, normalized by the corresponding genome size. 170

Figure 4-5. Inventory of genes encoding for enzymes involved in aerobic respiration, denitrification, and transcriptional regulation in the four most active MAGs. Genes that were present in a genome but were not differentially expressed (FDR > 0.05) were assigned the minimum absolute dynamic LFC value of 0.1. Genes that were not present in a particular MAG are indicated by the white squares. The dynamic changes in gene expression are shown as a gradient from light blue to dark purple, between minimum and maximum dynamic log₂ fold change in expression ($\sim 0.1 - 12$) (i.e. the sum of absolute value of log₂ fold change over the entire time series). 179

Figure 4-6. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the CHB metagenome (this study) and 9 published Chlorobi and Ignavibacteriae genomes from the phylogenetic tree. All genomes were greater than 90% complete with less than 2.5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 16,510 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: ‘Core’ (gene clusters shared between all genomes), ‘Shared’ (gene clusters present in CHB and closely related genomes Chlorobi bacterium OLB4 [JYPE01] and Ignavibacteriales UTCHB1[MWSW01]), ‘Unique’ (gene clusters present in only CHB, indicated on the outside of all rings), ‘Singletons’ (other GCs present in only one genome) and ‘Other’ (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices. 181

Figure 4-7. Response Groups 1-5 containing time-series gene expression profiles for the four most transcriptionally active MAGs in the MAD MA-AX microbiome. Expression profiles for differentially expressed genes related to inorganic nitrogen cycling were clustered with a nonparametric model-based method, DPGP, in each MAG. Gene expression response types for these MAGs were grouped based on similar response types. The y-axis represents the standardized log₂ fold change in gene expression over the time series and the x-axis is the sampling interval time in minutes. The shaded region in each plot represents the samples collected during the microaerobic period, followed by an anoxic period. The upper left corner of each plot denotes the MAG-specific subgroup within a particular response group (numbers circled in red). The number of genes clustered and statistical significance are denoted in the bottom right corner of each individual plot 184

Figure 4-8. Response Groups 6-10 containing time-series gene expression profiles for the four most transcriptionally active MAGs in the MAD MA-AX microbiome. Expression profiles for differentially expressed genes related to inorganic nitrogen cycling were clustered with a nonparametric model-based method, DPGP, in each MAG. Gene expression response types for these MAGs were grouped based on similar response types. The y-axis represents the standardized log₂ fold change in gene expression over the time series and the x-axis is the sampling interval time in minutes. The shaded region in each plot represents the samples collected during the microaerobic period, followed by an anoxic period. The upper left corner of each plot denotes the MAG-specific subgroup within a particular response group (numbers circled in red). The number of genes clustered and statistical significance are denoted in the bottom right corner of each individual plot. 190

Figure 4-9. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the AOB metagenome (this study) and 16 published Nitrosomonas genomes from the phylogenetic tree. All genomes were greater than 95% complete with less than 4% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 14,303 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: ‘Core genes’ (gene clusters shared between all genomes); four ‘Shared’ bins (gene clusters shared between AOB and Nitrosomonas europaea [AL954747], AOB and Nitrosomonas eutropha [NC_008344], shared between all three of the former genomes, and shared between all N. europaea lineage genomes indicated by the dark blue rings); ‘Unique’ (gene clusters present in only AOB, indicated on the outside of all rings); ‘Singletons’ (other GCs present in only one genome) and ‘Other’ (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices. 196

Figure 4-10. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the NOB metagenome (this study) and 7 published Nitrospira genomes in the phylogenetic

tree. All genomes were greater than 97% complete with less than 5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 12,155 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: 'Core genes' (gene clusters shared between all genomes), 'Shared' (gene clusters present in NOB and closely related genome *Nitrospira defluvii* [NC074355]), 'Unique' (gene clusters present in only NOB, indicated on the outside of all rings), 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices. 206

Figure 4-11. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the AMX metagenome (this study) and 5 published anammox genomes from the phylogenetic tree. All genomes were greater than 91% complete with less than 5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 7,706 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: 'Core genes' (gene clusters shared between all genomes), 'Shared' (gene clusters present in AMX and closely related genome *Ca. Brocadia fulgida* [LAQJ01]), 'Unique' (gene clusters present in only AMX, indicated on the outside of all rings), 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices. 214

Figure 5-1. WRRF overview with proposed modifications represented as dotted lines. Food waste codigestion is used to boost methane production in the anaerobic digesters (2), while increasing other nutrients in anaerobic digester effluent (3). Phosphorus (P) is released through WAS blending and mixed with digester filtrate and centrate prior to high-quality fertilizer production from struvite harvesting (4). Struvite harvesters precipitate a majority of P (with magnesium and ammonium); however, a high strength N effluent remains and is recycled back to the head of the plant. Since food waste codigestion is expected to increase N loads even further- we propose the addition of a micro-aerobic/anoxic bioreactor (5) following struvite harvesting to alleviate the N load to the mainstream. The simultaneous use of micro-aerobic mainstream BNR (6) reduces overall energy and carbon demand.....264

List of Supplementary Figures

Figure S2-1. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Candidatus Accumulibacter* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Ca. Accumulibacter* OTUs, not the whole community. 74

Figure S2-2. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Nitrosomonas* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Nitrosomonas* OTUs, not the whole community. 75

Figure S2-3. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Nitrospira* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Nitrospira* OTUs, not the whole community. 76

Figure S2-4. Neighbor-joining consensus tree generated from an alignment of published 16S rRNA sequences and sequences retrieved in this study. *Nitrobacter winogradskyi* was used to root the tree. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. The scale bar indicates a 4% sequence difference. Accession numbers are presented after the sequence names. 77

Figure S3-1. Total comammox assay (Ntsp-amoA 162F/359R) melting peaks for A) *Ca. N. nitrosa* amoA standard curve, B) *Ca. N. inopinata* amoA standard curve, and C) *Ca. N. nitrificans* amoA standard curve, D) L-SBR samples, E) P-SBR samples, F) P-CF samples, G) TRA samples, and H) NS samples. The dashed lines in panels A-C represent the standard deviation of melting temperatures for that specific standard. The dashed lines in panels D-H represent the minimum and maximum range of melting temperatures derived from all the standard curves. Triplicate data series are shown for each sample. 133

Figure S3-2. Agarose gel electrophoresis was performed using the total comammox (Ntsp-amoA 162F/359R) assay and included one replicate from each of the following samples: a non-template control (ntc), *Ca. N. nitrosa* amoA standard (nitro.std), *Ca. N. inopinata* amoA standard (inop.std), *Ca. N. nitrificans* amoA standard (nitrif.std), and one replicate from each of the 19 bioreactor samples. The expected amplicon size is 198 bp for this assay. **Figure S3-3.** Clade A comammox assay (coma-244f/659r) melting peaks for A) *Ca. N. nitrosa* amoA standard curve, B) *Ca. N. inopinata* amoA standard curve, C) *Ca. N. nitrificans* amoA standard curve, D) L-SBR samples, E) P-SBR samples, F) P-CF samples, G) TRA samples, and H) NS samples. The dashed lines in panels A-C represent the standard deviation of melting temperatures for that specific standard. The dashed lines in panels D-H represent the minimum and maximum range of melting temperatures derived from all the standard curves. Triplicate data series are shown for each sample. 134

Figure S3-4. Agarose gel electrophoresis was performed using the clade A comammox (comaA-244f/659r) assay and included one replicate from each of the following samples: a non-template control (ntc), *Ca. N. nitrosa* amoA standard (nitro.std), *Ca. N. inopinata* amoA standard (inop.std), *Ca. N. nitrificans* amoA standard (nitrif.std), and one replicate from each of the 19 bioreactor samples. The expected amplicon size for this assay is 415 bp. **Figure S3-5.** Clade B comammox assay (coma-244f/659r) melting peaks for A) *Nitrospira* sp CG24_E amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the standard deviation of melting temperatures for the standard. Triplicate data series are shown for each sample. 136

Figure S3-6. Agarose gel electrophoresis was performed using the clade B comammox (comaB-244f/659r) assay and included one replicate from each of the following samples: a non-template control (ntc),

Nitrospira sp. CG24_E *amoA* standard (CG24E), and one replicate from each of the 19 bioreactor samples. The expected amplicon size for this assay is 415 bp 138

Figure S3-7. Total comammox (Ntsp-*amoA* 162F/359R) gradient PCR assay results with *Ca. N. nitrosa* standard and two bioreactor samples (A) P_SBR-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 47.8 °C and 56.0 °C, where 48 °C is the original design annealing temperature. The expected amplicon size is 198 bp for this assay. Off-target products are observed at approximately 300 bp and 600 bp in the P_SBR-4 sample and at approximately 150 bp, 350 bp, 600 bp and 700 bp in the NS-4 sample, and appear to be independent of the total DNA concentrations tested in the reaction. 139

Figure S3-8. Clade A comammox (comaA-244f/659r) gradient PCR assay results with *Ca. N. nitrosa* standard and two bioreactor samples (A) P_CF-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 51.8 °C and 60.0 °C, where 52 °C is the original design annealing temperature. The expected amplicon size for this assay is 415 bp. Off-target products are observed at approximately 150 bp and 300 bp in the P_CF-4 sample and at approximately 250 bp and 300 bp in the NS-4 sample, and the amplification of these off-target products appear to be independent of the total DNA concentrations tested in the reaction. That is, with the exception of the P_CF-4 sample at the lowest total DNA concentration tested, where the 150 bp off-target product appears brighter than the target at the design annealing temperature, indicating the off-target product may be selectively amplified over the target in these conditions. 140

Figure S3-9. Clade B comammox (comaB-244f/659r) gradient PCR assay results with *Nitrospira* sp. CG-24E standard and two bioreactor samples (A) P_CF-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 51.8 °C and 60.0 °C, where 52 °C is the original design annealing temperature. The expected amplicon size for this assay is 415 bp. Multiple off-target products, especially one at 300 bp, appear at nearly every DNA concentration and temperature tested. 141

Figure S3-10. Comammox *Ca. N. nitrosa* assay (Nitrosa *amoA*-469F/812R) melting peaks after 45 cycles for A) *N. nitrosa amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample. 142

Figure S3-11. Comammox *Ca. N. nitrosa* assay (Nitrosa *amoA*-469F/812R) melting peaks after 30 cycles for A) *N. nitrosa amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample. 143

Figure S3-12. Comammox *Ca. N. inopinata* assay (Inopinata *amoA*-410F/815R) melting peaks after 45 cycles for A) *N. inopinata amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperatures for the standard. Triplicate data series are shown for each sample. 144

Figure S3-13. Comammox *Ca. N. inopinata* assay (Inopinata *amoA*-410F/815R) melting peaks after 30 cycles for A) *N. inopinata amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample. 145

Figure S3-14. Comammox *N. nitrificans* assay (Nitrificans *amoA*-463F/836R) melting peaks after 45 cycles for A) *N. nitrificans amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample. 146

- TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperatures for the standard. Triplicate data series are shown for each sample. 146
- Figure S3-15.** Comammox *N. nitrificans* assay (Nitrificans amoA-463F/836R) melting peaks after 30 cycles for A) *N. nitrificans* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample. 147
- Figure S3-16.** A qPCR assay with 30 cycles was generated to test each new primer set targeting A) *Ca. N. nitrosa*, B) *N. inopinata*, and C) *N. nitrificans* in one DNA sample from each bioreactor (lanes 2-20) with its corresponding non-template control (NTC, lane 22) and standard (lane 23). The expected band was present with amoA template from each external standard and some bioreactor samples. No off-target products were observed in the non-template control (NTC), bioreactor samples, or standards. Lanes 1, 21, and 24 correspond to 100 bp DNA size marker. 148
- Figure S3-17.** Agarose gel electrophoresis of qPCR products amplified after 45 cycles with the primers-set designed in this study targeting the amoA gene of *N. nitrosa*. The first and last lanes correspond to 100 bp DNA size marker. 149
- Figure S3-18.** A PCR assay was generated to test each new primer set targeting *Ca. N. nitrosa*, *N. inopinata*, and *N. nitrificans* with its corresponding standard, a non-template control, and one DNA sample from each bioreactor. The expected band was present with amoA template from each external standard (*Ca. N. nitrosa* (nitro.), *Ca. N. inopinata* (inop.), and *Ca. N. nitrificans* (nitrif.)) and some bioreactor samples. No off-target products were observed in the non-template control (ntc), bioreactor samples, or standards. The first and last lanes correspond to 100 bp DNA size marker. 150
- Figure S3-19.** Cross-hybridization and primer dimers were not observed for the newly designed primer sets when combined in reaction with amoA template from each external standard (*Ca. N. nitrosa* (nitro.), *Ca. N. inopinata* (inop.), and *Ca. N. nitrificans* (nitrif.)) and non-template control (ntc). The first and last lanes correspond to 100 bp DNA size marker. 151
- Figure S3-20.** Percent relative abundance of comammox *Ca. Nitrospira* amoA genes to total bacteria (amoA gene copies/ 16S rRNA gene copies). Note the y-axis is not linear and is separated into three sections to show relative abundances below 0.30%, between 0.30 and 3.0%, and between 3.0 to 10%. 152
- Figure S4-1.** Amino acid identity (AAI) and bit score distribution of CHB versus *Chlorobi bacterium* OLB4 253
- Figure S4-2.** Amino acid identity (AAI) and bit score distribution of CHB versus *Ignavibacteriales* UTCHB1 254
- Figure S4-3.** Amino acid identity (AAI) and bit score distribution of AOB versus *Nitrosomonas europaea* 255
- Figure S4-4.** Amino acid identity (AAI) and bit score distribution of AOB versus *Nitrosomonas* OLB2.. 256
- Figure S4-5.** Amino acid identity (AAI) and bit score distribution of NOB versus *Nitrospira defluvii* .. 257
- Figure S4-6.** Amino acid identity (AAI) and bit score distribution of NOB versus *Nitrospira* OLB3..... 258
- Figure S4-7.** Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia fulgida* 259

Figure S4-8. Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia sinica* OLB1 260

Figure S4-9. Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia* UTAMX2261

Figure S4-10. Amino acid identity (AAI) and bit score distribution of *Ca. Brocadia* UTAMX2 versus *Ca. Brocadia fulgida* 262

List of Tables

Table 2-1. Best-fit Monod kinetic parameters for ammonia oxidation and phosphorus uptake from the full-scale plant and from the pilot-scale plant after stable low-DO operation.	42
Table 3-1. Bioreactor sample characteristics.....	90
Table 3-2. Primers used for qPCR.....	94
Table 3-3. qPCR performance with three distinct comammox <i>amoA</i> standard templates. Average parameters are presented with standard deviation. PCR efficiency is reported for the linear range of the standard calibration curve.	97
Table 4-1. Reactor properties, chemical data statistics, and performance summary.....	162
Table 4-2. Genome statistics of the top 16 MAGs recovered from the MAD MA-AX community. MAGs with the greatest relative expression are shaded green.	165

List of Supplementary Tables

Table S2-1. Average operational parameters for the pilot- and full-scale plants.....	71
Table S2-2. 16S rRNA-targeted oligonucleotide probes used in this study	72
Table S2-3. Energy savings summary table, including modeled basin DO, calculated total energy usage, percent energy savings and change in annual energy costs based on \$0.0835/kWh. Shaded rows represent the current scenario and the low-DO scenario that corresponds to the stable low-DO operation of the pilot plant.	73
Table S3-1. qPCR standard curves: slope, y-intercept, correlation coefficient, and PCR efficiency calculated from the approximate linear range for each assay included in this study.	125
Table S3-2. Total comammox primer mismatch table with comammox standard template <i>amoA</i> used in this study	126
Table S3-3. Total comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive.....	127
Table S3-4. Clade A comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive	128
Table S3-5. Clade B comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive.....	129
Table S3-6. <i>Ca. N. nitrosa</i> comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive	130
Table S3-7. <i>Ca. N. inopinata</i> comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive.....	131
Table S3-8. <i>Ca. N. nitrificans</i> comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive.....	132
Table S4-1. Samples selected for metagenomic sequencing, respective total inorganic nitrogen (TIN) performance, and 16S rRNA sequencing relative abundance from key functional groups.	245
Table S4-2. Metagenomic DNA read summary and statistics.....	246
Table S4-3. Metatranscriptomic cDNA read summary and statistics.....	247
Table S4-4. Genome statistics of the highly complete/low redundancy MAGs	248
Table S4-5. Summary of CHB gene expression clustering.....	249
Table S4-6. Summary of AOB gene expression clustering	250
Table S4-7. Summary of NOB gene expression clustering	251
Table S4-8. Summary of AMX gene expression clustering	252

Chapter 1

General Introduction

Global significance of wastewater treatment

Nutrients, such as nitrogen (N) and phosphorus (P), are essential for life, yet are detrimental to ecosystems when present in excess. Human activities introduce a continuous disruption of the natural global N and P cycles. Thus, tracking and improving our understanding of anthropogenic N and P, while undeniably difficult, is essential as N and P play a role in all living systems and many environmental problems [1-3].

In natural systems, N is present in the atmosphere as nitrogen gas, nitrogen oxides, and ammonia, and in water as dissolved inorganic nitrogen, dissolved organic nitrogen, and particulate organic nitrogen compounds. Prior to pre-industrial times, fixation of nitrogen gas into biologically available N was primarily driven by highly specialized microorganisms. However, human production of additional nitrogen nutrients has doubled the amount of nitrogen fixation [4], where most of this fixed N is used as fertilizers in agriculture and N is either volatilized into greenhouse gas (e.g. nitrous oxide, N_2O), transported to rivers and coastal areas through surface water runoff, or infiltrated into groundwater [4]. All-in-all, since the late-19th century, inputs of biologically available N into the environment have increased more than tenfold [1, 5].

Unlike N, P does not include a stable atmospheric gas phase; thus, in natural systems, P primarily exists as soluble phosphorus (orthophosphate or dissolved organic phosphorus) or particulate phosphorus (adsorbed phosphate, insoluble P minerals, particulate organic P). Prior to anthropogenic influence, availability of P was restricted to natural chemical weathering with transport limited to surface and subsurface runoff; thus, ecosystems depended on transport of P through the aqueous phase [2, 6]. The modern P cycle, on the other hand, has been increasingly manipulated by the

production and transport of fertilizers for agriculture, livestock waste, industrial applications, deforestation and increased erosion, as well as human waste [6]. Collectively, these human-derived activities have doubled the net input of dissolved P into the environment [2].

High loads N and P into surface waters remain a significant contributor to eutrophication and oxygen depletion or “dead zones” in freshwater and coastal regions such as the Gulf of Mexico [7]. In the United States (U.S.) alone, damage caused by eutrophication costs approximately \$2.2 billion annually [8]. In developing countries, approximately 90 percent of all wastewater is untreated and discharged directly into rivers, lakes, and oceans alike [9]. From this, an estimated 94,595 mi² (245,000 km²) of marine ecosystems have already been affected worldwide, which not only damages fisheries but also the livelihood of the communities that depend on them [9].

In the U.S., WWTP effluent water quality has greatly improved since the early 1970s thanks to treatment plant upgrades made to meet requirements of the Clean Water Act (CWA) [10, 11]. Existing WWTPs are often equipped with a biological nutrient removal (BNR) process, which take advantage of the metabolism of microorganisms to effectively remove or transform nutrients into less-toxic compounds. While BNR provides many advantages to a WWTP, pumping oxygen to this process is often energy-intensive and costly. As humanity also faces a looming energy-crisis, wastewater treatment facilities are also working to become energy-independent while maintaining effective effluent water quality [12]. **This goal can be realized, in part, through a comprehensive understanding of the mechanisms behind N and P cycling in BNR**

processes operated with less oxygen input (i.e. ‘low dissolved oxygen [low DO]’ conditions), as discussed in Chapter 2.

Our improved understanding of the mechanisms and microorganisms driving low DO BNR will be a fundamental advancement towards energy efficient and effective treatment. In the U.S., this is especially important as more plants must meet increasingly stringent total nitrogen (TN) and P permit requirements [13]. In developing countries, an energy-independent wastewater treatment facility may allow for greater implementation, with the potential to improve health, protect natural resources, and provide employment around the world [9].

Nitrogen removal

Nitrogen (N) has one of the most intricate cycles of all the major elements, as it exists in eight oxidation states and is transformed through several mechanisms within a diverse set of microorganisms [4, 14] (Figure 1-1). Over the past three decades, scientists have redefined our understanding of the microbial nitrogen (N) transformation network— with the discovery of new pathways and microorganisms involved in global N cycle. The breakthrough discoveries of anaerobic ammonium oxidizing bacteria (anammox) [15], ammonia oxidizing archaea (AOA) [16], complete ammonia oxidation (comammox) by one organism [17, 18], and the ability for N-cycling organisms to use alternative substrates [19, 20] or simultaneous pathways [21, 22] reveals how incredibly versatile and complex these microbial metabolisms are. In the past decade, these discoveries have transformed into exciting new opportunities for N removal from wastewater.

In conventional BNR systems *without* TN discharge limitations, ammoniacal nitrogen (ammonium $[\text{NH}_4^+]$ + ammonia $[\text{NH}_3]$) is removed through the microbially mediated oxidation of ammonium (NH_4^+) via nitrite (NO_2^-) to nitrate (NO_3^-), also known as nitrification (Steps 1 - 4 in Figure 1-1). If TN removal is required, wastewater treatment facilities will additionally enrich for microorganisms that carry out the reduction of NO_2^- and NO_3^- to atmospheric nitrogen (N_2), also known as denitrification (Steps 5 - 8 in Figure 1-1).

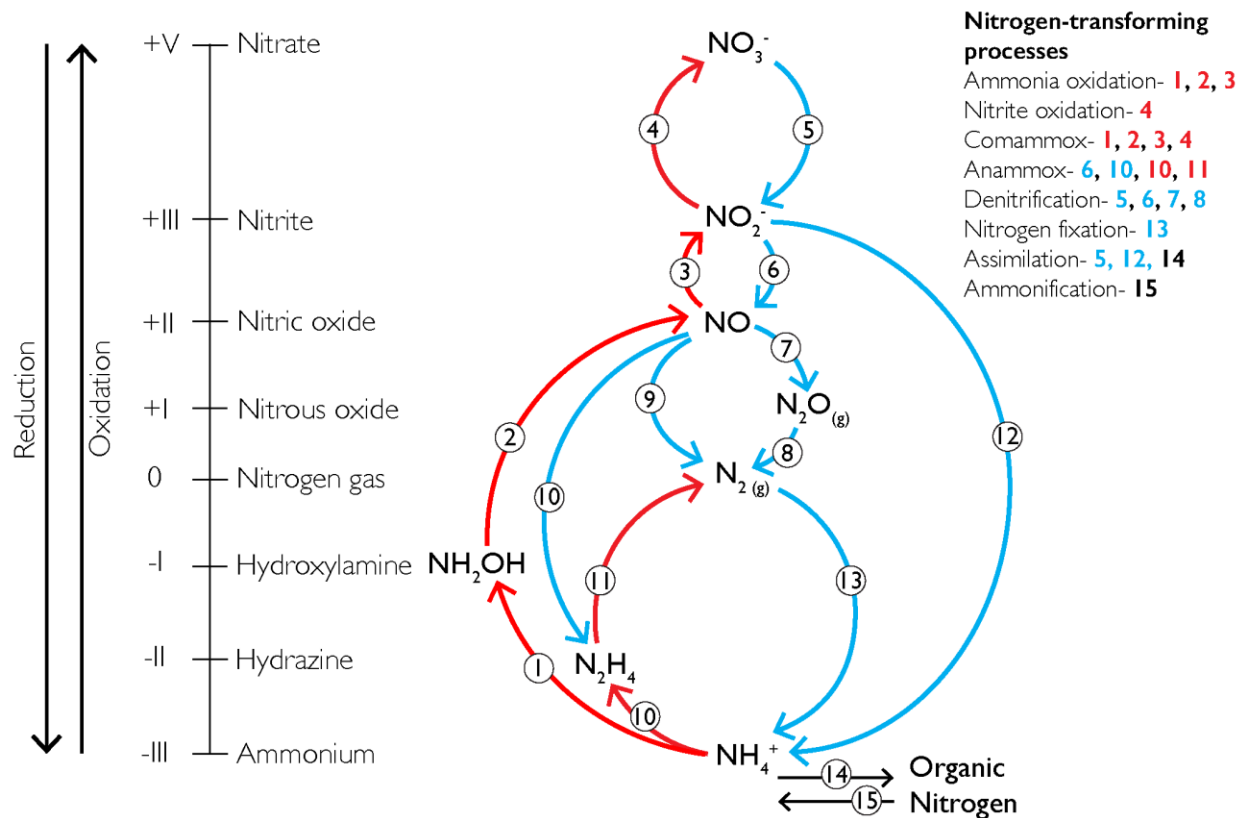
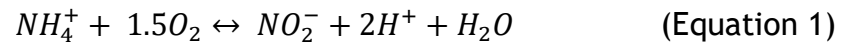


Figure 1-1. Fifteen transformations among eight key nitrogen redox states (adapted from Kuypers et al. [14]). Red arrows represent an oxidation reaction that requires oxygen or another electron acceptor; blue arrows represent a reduction reaction, where organic carbon or other available electron donors are needed to complete the pathway. While nitrogen-transforming processes (upper-right) describe specific redox pathways; the microorganisms belonging to key functional groups are actually metabolically versatile and may contain multiple capabilities and/or may perform pathways simultaneously.

Nitrification

Nitrification is the fundamental link in the global biogeochemical nitrogen cycle that is exclusively carried out by prokaryotes [16, 23, 24]. Nitrifying microorganisms are chemolithoautotrophic aerobes that use fixed carbon dioxide as their carbon source and oxygen as their terminal electron acceptor when transforming NH_4^+ via NO_2^- to nitrate NO_3^- (Steps 1 - 4 in Figure 1-1) [25]. Ammonium oxidizing bacteria (AOB) and archaea (AOA) are known to complete the first step of nitrification (Steps 1 - 3 in Figure 1-1), the oxidation of ammonium to nitrite (Equation 1) [25].



$$\Delta G^{0'} = -274.7 \text{ kJ mol}^{-1}$$

Ammonium oxidizing microorganisms belong to genera with the prefix 'Nitroso-' [25]. Specifically, the most frequently detected AOB primarily belong to the genera *Nitrospira* and *Nitrosomonas* [26, 27]. AOA, also carry out ammonium oxidation in some environments, and primarily belong to the genera *Nitrosopumilus* and *Nitrososphaera* [28].

Nitrite oxidizing bacteria (NOB) are known to carry out the second step in nitrification (Figure 1-1: process 4) by oxidizing nitrite to nitrate (Equation 2) and belong to genera with the prefix 'Nitro-' [25]. NOB populations belong primarily to the genera *Nitrobacter* and *Nitrospira* [25, 29-31].

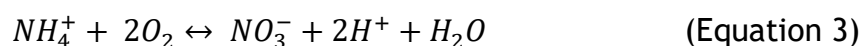


$$\Delta G^{0'} = -74.1 \text{ kJ mol}^{-1}$$

Both ammonium oxidation and nitrite oxidation are energetically favorable reactions, as indicated by the negative Gibbs free energy.

Since the late 1800s, nitrification had been described as a division of labor between the two phylogenetically distinct and specialized chemolithoautotrophic organisms, AOB/AOA and NOB [32]. The perception that two organisms must work together to carry out complete nitrification was not challenged for over a century, until Costa et al. hypothesized that a single nitrifying bacterium combining ammonia oxidation and nitrite oxidation, ‘comammox’ (complete ammonium oxidation), should exist in nature [33].

Complete ammonium oxidation (Equation 3) yields more energy than either of the two individual nitrification steps, which is why the separation of nitrification steps between two functional groups of microorganisms had perplexed many scientists [17].



$$\Delta G^{0'} = -348.9 \text{ kJ mol}^{-1}$$

Nearly one decade after the complete ammonium oxidation process was described by Costa et al., in late 2015, two separate research groups discovered, cultivated, and characterized the ‘missing’ comammox organism from an aquaculture system and a deep oil exploration well [17, 18]. Intriguingly, these comammox organisms belong to the genus *Nitrospira*, a NOB believed to only carry out the second step of nitrification. Genome enabled research has revealed that comammox *Nitrospira* have the machinery to carry out both steps of nitrification. A recent study [34] suggests that comammox may not be significant to conventional biological nutrient removal, **we provide**

experimental evidence of comammox in energy-efficient BNR systems operated with low DO conditions (Chapter 3) [35].

Denitrification

Wastewater treatment plants that have a TN requirement (i.e. they must remove further reduce nitrogen in their effluent through denitrification) take advantage of the heterotrophic reduction of nitrate to N_2 [4]. The identity of these organisms and details about dominant populations are less clear because the ability to denitrify is widespread among organisms and populations seem to vary between plants with different influent properties and operational conditions [36]. However, in many plants, the required organic carbon compounds to complete denitrification (e.g. volatile fatty acids such as acetate, propionate, etc.) are a limiting factor.

Studies on nitrification have established that nitrification and denitrification can occur concurrently in the same reactor with low DO; this phenomenon is referred to as simultaneous nitrification and denitrification (SND) [37, 38]. When partial nitrification (NH_3 to NO_2^-) occurs and nitrite is directly reduced to N_2 , it is termed SND via the nitrite shunt (Steps 1 - 3 and 6 - 8 in Figure 1-1) [39, 40]. SND via the nitrite shunt may not only save energy (through the low DO operation) but also save 40% of the organic matter requirement for denitrification when compared to typical nitrification and denitrification processes [41].

Anammox

Another recently discovered N pathway, anaerobic ammonium oxidation (anammox), is already utilized as an energy-saving option in wastewater treatment. Anammox

microorganisms do not require oxygen or organic carbon as they convert NH_4^+ and NO_2^- directly to N_2 (Steps 6, 10 - 11 in Figure 1-1) [15]. The anammox process is viewed as a ‘shortcut’ to the classic biogeochemical nitrogen cycle. Yet, the slow growth rate and sensitivity to low temperature and high oxygen concentrations prevents anammox from becoming naturally enriched in conventional activated sludge systems [42]. **In Chapter 4, we discuss the operation of and the microbial community (including anammox bacteria) within a cyclic micro-aerobic (i.e. low DO) and anoxic sequencing batch reactor designed to treat the ammonia occurring in a wastewater treatment sidestream.**

Biological phosphorus removal

The enhanced biological phosphorus removal (EBPR) process serves as an alternative to phosphorus removal using chemical precipitation methods. Chemical precipitation is not always a favorable alternative for phosphorus (P) removal because it relies on chemicals that often have unstable prices. Further, chemically removed P is difficult to recover in downstream processes.

EBPR is primarily performed by polyphosphate accumulating organisms (PAO), with the ability to accumulate and store P intracellularly. EBPR processes typically begin with an anaerobic zone, where PAO take short-chained volatile fatty acids (VFA) such as acetate or propionate into their cells and store them as poly- β -hydroxyalkanoate (PHA) [52]. Intracellular polyphosphate (polyP) and glycogen are degraded to help PAO consume VFA. When polyP is degraded during this process, phosphorus is released from the cells [52]. Even though phosphorus release in the anaerobic zone of wastewater treatment seems counterproductive to overall phosphorus removal during wastewater

treatment, the release of P is essential and allows for increased phosphorus uptake in the presence of oxygen or other terminal electron acceptors. In the presence of oxygen, PAO use previously stored PHA as a source of carbon and energy in order to take phosphorus into their cells and replenish intracellular polyP supplies [52]. The anoxic uptake of phosphorus has also been observed in both laboratory and full-scale EBPR systems (where PAO use nitrite or nitrate is used as the electron acceptor rather than oxygen).

Studies of EBPR plants around the world have shown that *Candidatus* 'Accumulibacter phosphatis' (*Ca. Accumulibacter*) is the most prevalent PAO in WWTPs [29]. *Ca. Accumulibacter* populations are diverse, not only morphologically but also genetically [56-59] and have been divided into two main types (type I and type II) made up of several clades [60, 61]. One study demonstrated that a *Ca. Accumulibacter* type I clade was able to take in P using nitrate as a terminal electron acceptor, while some type II *Ca. Accumulibacter* clades rely strictly on oxygen for P uptake [54, 55]. The population of PAO with the ability to take up P under anoxic conditions are referred to as denitrifying polyphosphate accumulating organisms (DPAO).

The ability for PAO to use either nitrite or nitrate instead of oxygen for P uptake may be a key factor for energy saving EBPR processes; however, PAO need fermentation products for P release which may be a limiting factor for many wastewater facilities. However, some studies have suggested that an organism belonging to the genus *Tetrasphaera* may be a key PAO to overcome limitations, with its unique ability to perform fermentation, denitrification, and polyphosphate accumulation [63, 64].

Energy requirement for biological nutrient removal

Energy independence has been an expanding theme in WRRF research [12, 65-68]. Based on the stoichiometry, complete nitrification requires two moles of oxygen to oxidize every mole of ammonium; thus, it is not surprising that BNR is typically operated with extensive aeration and is energy intensive [69]. However, over the past 30 years, N removal via nitrification and denitrification has been demonstrated successfully with low DO concentrations ($< 1.0 \text{ O}_2/\text{L}$) from lab-scale to full-scale (Figure 1-2). While nitrification has been widely reported at low DO, the effect of low DO on EBPR is less clear. Therefore, in Chapter 2, we investigated, at the pilot-scale level, the effect of using minimal aeration on the performance of an EBPR process.

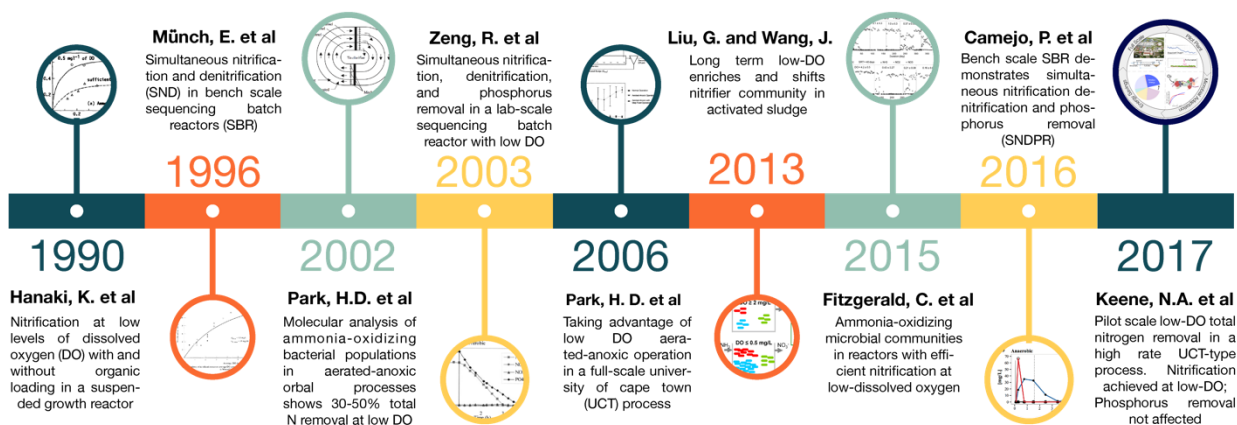


Figure 1-2. Evolving research of low oxygen biological nitrogen removal shows successful performance without a consensus identity to the key microbial players. [70, 73-79]

Although there are several alternative strategies to increase aeration efficiency within secondary treatment of a WWTP (e.g., replacing existing blowers with increased efficiency blowers, installing fine bubble membrane diffusers, etc.), many of these upgrades can be costly. For some WWTPs, operating with low DO could be a simple step towards achieving additional energy savings. For other facilities, upgrades to existing infrastructure may be necessary prior to implementing a low DO BNR. Ultimately, we calculated that once a full-scale BNR process is converted to low DO operation, a full-scale WWTP could save nearly 33% in total electricity, or approximately \$300,000 per year, without having to upgrade any existing inefficient equipment (Chapter 2) [70].

References

1. Galloway, J.N., et al., *Nitrogen footprints: past, present and future*. Environmental Research Letters, 2014. **9**(11): p. 115003.
2. Filippelli, G., *The global phosphorus cycle: Past, present, and future*. Elements, 2008. **4**(2): p. 89-95.
3. Galloway, J.N., et al., *Nitrogen cycles: past, present, and future*. Biogeochemistry, 2004. **70**(2): p. 153-226.
4. Ward, B., *The Global Nitrogen Cycle*, in *Fundamentals of Geobiology*. 2013, John Wiley & Sons, Ltd. p. 36-48.
5. Wang, J.S., et al., *Emissions credits: opportunity to promote integrated nitrogen management in the wastewater sector*. Environ Sci Technol, 2011. **45**(15): p. 6239-46.
6. McDowell, R.W. and A. Sharpley, *Approximating phosphorus release from soils to surface runoff and subsurface drainage*. Journal Of Environmental Quality, 2001. **30**(2): p. 508-520.
7. Rabalais, N.N., R.E. Turner, and W.J. Wiseman, *Gulf of Mexico hypoxia, aka "The dead zone"*. Annual Review of Ecology and Systematics, 2002. **33**: p. 235-263.
8. Dodds, W.K., *Eutrophication of U.S. freshwaters: analysis of potential economic damages*. Environmental Science and Technology, 2009.
9. UNEP, U.-H., GRID-Arendal, *Sick Water? The Central Role of Wastewater Management in Sustainable Development*. 2010.
10. Systems, C.f.S. *U.S. Wastewater Treatment*. 2014; Available from: http://css.snre.umich.edu/css_doc/CSS04-14.pdf.

11. Drury, B., E. Rosi-Marshall, and J.J. Kelly, *Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers*. Appl Environ Microbiol, 2013. **79**(6): p. 1897-905.
12. Remy, C., M. Boulestreau, and B. Lesjean, *Proof of concept for a new energy-positive wastewater treatment scheme*. Water Sci Technol, 2014. **70**(10): p. 1709-16.
13. EPA, U., *Clean Watersheds Needs Survey 2008 Report to Congress*, U.S.E.P. Agency, Editor. 2008.
14. Kuypers, M.M.M., H.K. Marchant, and B. Kartal, *The microbial nitrogen-cycling network*. Nat Rev Microbiol, 2018. **16**(5): p. 263-276.
15. van de Graaf, A.A., et al., *Anaerobic oxidation of ammonium is a biologically mediated process*. Applied and Environmental Microbiology, 1995. **61**(4): p. 1246-51.
16. Konneke, M., et al., *Isolation of an autotrophic ammonia-oxidizing marine archaeon*. Nature, 2005. **437**(7058): p. 543-6.
17. Daims, H., et al., *Complete nitrification by Nitrospira bacteria*. Nature, 2015. **528**(7583): p. 504-9.
18. van Kessel, M.A., et al., *Complete nitrification by a single microorganism*. Nature, 2015. **528**(7583): p. 555-9.
19. Daims, H., S. Lucker, and M. Wagner, *A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria*. Trends Microbiol, 2016. **24**(9): p. 699-712.
20. Fussel, J., et al., *Adaptability as the key to success for the ubiquitous marine nitrite oxidizer Nitrococcus*. Sci Adv, 2017. **3**(11): p. e1700807.
21. Stein, L.Y. and M.G. Klotz, *The nitrogen cycle*. Curr Biol, 2016. **26**(3): p. R94-8.

22. Yan, Y., et al., *Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated Pseudomonas stutzeri A1501*. Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7564-9.
23. Arrigo, K.R., *Marine microorganisms and global nutrient cycles*. Nature, 2005. **437**(7057): p. 349-55.
24. Bock, E. and M. Wagner, *Oxidation of Inorganic Nitrogen Compounds as an Energy Source*. 2006: p. 457-495.
25. Rittman, B.E. and P.L. McCarty, *Environmental Biotechnology: Principles and Application*. 2001, New York, NY: McGraw-Hill
26. Nielsen, P.H., et al., *A conceptual ecosystem model of microbial communities in enhanced biological phosphorus removal plants*. Water Res, 2010. **44**(17): p. 5070-88.
27. Ge, S., et al., *Detection of nitrifiers and evaluation of partial nitrification for wastewater treatment: A review*. Chemosphere, 2015. **140**: p. 85-98.
28. Monteiro, M., J. Seneca, and C. Magalhaes, *The history of aerobic ammonia oxidizers: from the first discoveries to today*. J Microbiol, 2014. **52**(7): p. 537-47.
29. Seviour, R.J., Nielsen, Per H, *Microbial Communities in Activated Sludge Plants*, in *Microbial Ecology of Activated Sludge*, R.J. Seviour, Nielsen, Per H, Editor. 2010, IWA Publishing: London.
30. Lucker, S., et al., *Nitrotoga-like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants*. ISME J, 2015. **9**(3): p. 708-20.
31. Saunders, A.M., et al., *The activated sludge ecosystem contains a core community of abundant organisms*. ISME J, 2016. **10**(1): p. 11-20.

32. Brock, T.D., *Milestones in Microbiology 1546 to 1940*. American Society Mic Series. 1999: ASM Press.
33. Costa, E., J. Perez, and J.U. Kreft, *Why is metabolic labour divided in nitrification?* Trends Microbiol, 2006. **14**(5): p. 213-9.
34. Chao, Y., et al., *Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach*. Appl Microbiol Biotechnol, 2016. **100**(18): p. 8225-37.
35. Beach, N.K. and D.R. Noguera, *Design and Assessment of Species-Level qPCR Primers Targeting Comammox*. Frontiers in Microbiology, 2019. **10**(36).
36. Seviour, R.J., *An Overview of the Microbes in Activated Sludge*, in *Microbial Ecology of Activated Sludge*, R.J. Seviour, Nielsen, Per H, Editor. 2010, IWA Publishing: London.
37. Bertanza, G., *Simultaneous nitrification-denitrification process in extended aeration plants: Pilot and real scale experiences*. Water Science And Technology, 1997. **35**(6): p. 53-61.
38. Helmer, C. and S. Kunst, *Simultaneous nitrification/denitrification in an aerobic biofilm system*. Water Science And Technology, 1998. **37**(4-5): p. 183-187.
39. Surmacz-Gorska, J., A. Cichon, and K. Miksch, *Nitrogen removal from wastewater with high ammonia nitrogen concentration via shorter nitrification and denitrification*. Water Science And Technology, 1997. **36**(10): p. 73-78.
40. Yoo, K., et al., *Nitrogen removal from synthetic wastewater by simultaneous nitrification and denitrification (SND) via nitrite in an intermittently-aerated reactor*. Water Research, 1999. **33**(1): p. 145-154.
41. Turk, O. and D.S. Mavinic, *Preliminary assessment of a shortcut in nitrogen removal from wastewater*. Canadian Journal of Civil Engineering, 1986. **13**(6): p. 600-605.

42. He, S., et al., *The Treatment Performance and the Bacteria Preservation of Anammox: A Review*. Water, Air, & Soil Pollution, 2015. **226**(5).
43. Bartossek, R., et al., *Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context*. Environmental microbiology, 2010. **12**(4): p. 1075-1088.
44. Caranto, J.D. and K.M. Lancaster, *Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase*. Proceedings of the National Academy of Sciences, 2017. **114**(31): p. 8217-8222.
45. Zorz, J.K., et al., *Comparative Proteomics of Three Species of Ammonia-Oxidizing Bacteria*. Frontiers in Microbiology, 2018. **9**.
46. Camejo, P.Y., et al., *Genome-Enabled Insights into the Ecophysiology of the Comammox Bacterium "Candidatus Nitrospira nitrosa"*. mSystems, 2017. **2**(5).
47. Lawson, C.E., et al., *Metabolic network analysis reveals microbial community interactions in anammox granules*. Nature Communications, 2017. **8**: p. 15416.
48. Strous, M., et al., *Deciphering the evolution and metabolism of an anammox bacterium from a community genome*. Nature, 2006. **440**(7085): p. 790-4.
49. Kraft, B., et al., *Nitrogen cycling. The environmental controls that govern the end product of bacterial nitrate respiration*. Science, 2014. **345**(6197): p. 676-9.
50. Lam, P. and M.M. Kuypers, *Microbial nitrogen cycling processes in oxygen minimum zones*. Ann Rev Mar Sci, 2011. **3**: p. 317-45.
51. Ravishankara, A.R., J.S. Daniel, and R.W. Portmann, *Nitrous oxide (N₂O): the dominant ozone-depleting substance emitted in the 21st century*. Science, 2009. **326**(5949): p. 123-5.

52. McMahon, K.D.H., Shaomei; Oehmen, Adrien, *The Microbiology of Phosphorus Removal*, in *Microbial Ecology of Activated Sludge*, P.H.N. Robert J Seviour, Editor. 2010, IWA Publishing: London.
53. Meinhold, J., *Characterization of the denitrifying fraction of phosphate accumulating organisms in biological phosphate removal*. 1999.
54. Flowers, J.J., et al., *Denitrification capabilities of two biological phosphorus removal sludges dominated by different 'Candidatus Accumulibacter' clades*. Environmental Microbiology Reports, 2009. 1(6): p. 583-588.
55. Oehmen, A., et al., *Assessing the abundance and activity of denitrifying polyphosphate accumulating organisms through molecular and chemical techniques*. Water Sci Technol, 2010. 61(8): p. 2061-8.
56. Carvalho, G., et al., *Denitrifying phosphorus removal: Linking the process performance with the microbial community structure*. Water Research, 2007. 41(19): p. 4383-4396.
57. Garcia Martin, H., et al., *Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities*. Nature Biotechnology, 2006. 24(10): p. 1263-1269.
58. He, S., A. Gu, and K. McMahon, *Progress Toward Understanding the Distribution of Accumulibacter Among Full-Scale Enhanced Biological Phosphorus Removal Systems*. Microbial Ecology, 2008. 55(2): p. 229-236.
59. Kong, Y., J.L. Nielsen, and P.H. Nielsen, *Microautoradiographic Study of Rhodocyclus-Related Polyphosphate-Accumulating Bacteria in Full-Scale Enhanced Biological Phosphorus Removal Plants*. Applied and Environmental Microbiology, 2004. 70(9): p. 5383-5390.

60. He, S., D.L. Gall, and K.D. McMahon, "*Candidatus Accumulibacter*" population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Appl Environ Microbiol*, 2007. **73**(18): p. 5865-74.
61. Peterson, S.B., et al., *Environmental distribution and population biology of Candidatus Accumulibacter, a primary agent of Biological Phosphorus Removal*. *Environmental microbiology*, 2008. **10**(10): p. 2692-2703.
62. Kim, J.M., et al., *Characterization of the denitrification-associated phosphorus uptake properties of "Candidatus Accumulibacter phosphatis" clades in sludge subjected to enhanced biological phosphorus removal*. *Appl Environ Microbiol*, 2013. **79**(6): p. 1969-79.
63. Kristiansen, R., et al., *A metabolic model for members of the genus Tetrasphaera involved in enhanced biological phosphorus removal*. *ISME Journal*, 2013. **7**(3): p. 543-554.
64. Mielczarek, A.T., et al., *Population dynamics of bacteria involved in enhanced biological phosphorus removal in Danish wastewater treatment plants*. *Water Research*, 2013. **47**(4): p. 1529-1544.
65. Nowak, O., P. Enderle, and P. Varbanov, *Ways to optimize the energy balance of municipal wastewater systems: lessons learned from Austrian applications*. *Journal of Cleaner Production*, 2015. **88**: p. 125-131.
66. Jenicek, P., et al., *Energy self-sufficient sewage wastewater treatment plants: is optimized anaerobic sludge digestion the key?* *Water Sci Technol*, 2013. **68**(8): p. 1739-44.
67. Gao, H., Y.D. Scherson, and G.F. Wells, *Towards energy neutral wastewater treatment: methodology and state of the art*. *Environ Sci Process Impacts*, 2014. **16**(6): p. 1223-46.

68. Zaborowska, E., K. Czerwionka, and J. Makinia, *Strategies for achieving energy neutrality in biological nutrient removal systems - a case study of the Slupsk WWTP (northern Poland)*. Water Science and Technology, 2016.
69. Hamilton, G., et al., *Driving energy efficiency in the US Water & Waste water industry by focusing on operating and maintenance cost reductions*. ACEEE Summer Study on Energy Efficiency in Industry, Washington, DC, 2009.
70. Keene, N.A., et al., *Pilot plant demonstration of stable and efficient high rate biological nutrient removal with low dissolved oxygen conditions*. Water Res, 2017. **121**: p. 72-85.
71. Tchobanoglous, G.B., Franklin L.; Stensel, H. David, *Wastewater Engineering: Treatment and Reuse*. 4 ed, ed. M.E. Inc. 2003, New York, NY: McGraw-Hill.
72. Park, H.D. and D.R. Noguera, *Evaluating the effect of dissolved oxygen on ammonia-oxidizing bacterial communities in activated sludge*. Water Res, 2004. **38**(14-15): p. 3275-86.
73. Hanaki, K.W., C.; Ohgaki, S., *Nitrification at low levels of dissolved oxygen with and without organic loading in a suspended growth reactor*. Water Research, 1989. **24**(3): p. 297-302.
74. V. Münch, E.L., Paul; Keller, Jürg, *Simultaneous Nitrification and Denitrification in Bench-Scale Sequencing Batch Reactors*. Water Research, 1996. **30**(2): p. 277-284.
75. Zeng, R.J., et al., *Simultaneous nitrification, denitrification, and phosphorus removal in a lab-scale sequencing batch reactor*. Biotechnol Bioeng, 2003. **84**(2): p. 170-8.
76. Park, H.D., et al., *Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors*. Applied and Environmental Microbiology, 2006. **72**(8): p. 5643-5647.

77. Liu, G. and J. Wang, *Long-term low DO enriches and shifts nitrifier community in activated sludge*. Environ Sci Technol, 2013. **47**(10): p. 5109-17.
78. Fitzgerald, C.M., et al., *Ammonia-oxidizing microbial communities in reactors with efficient nitrification at low-dissolved oxygen*. Water Res, 2015. **70**: p. 38-51.
79. Camejo, P.Y., et al., *Candidatus Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors*. Water Res, 2016. **102**: p. 125-137.
80. Lebedeva, E.V., et al., *Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon "Ca. Nitrosotenuis uzonensis" representing a clade globally distributed in thermal habitats*. PLoS One, 2013. **8**(11): p. e80835.
81. Henze, M., *Wastewater treatment : biological and chemical processes*. Environmental engineering. 1995, Berlin ; New York: Springer-Verlag. 383 pages.
82. Lu, H., K. Chandran, and D. Stensel, *Microbial ecology of denitrification in biological wastewater treatment*. Water Res, 2014. **64**: p. 237-54.
83. Scarborough, M.J., *Production Of Beneficial Chemicals From Renewable Feedstocks With Anaerobic Microbiomes*. 2018, University of Wisconsin- Madison.
84. Zhang, C., et al., *Reviewing the anaerobic digestion of food waste for biogas production*. Renewable and Sustainable Energy Reviews, 2014. **38**: p. 383-392.
85. Nghiem, L.D., et al., *Full scale co-digestion of wastewater sludge and food waste: Bottlenecks and possibilities*. Renewable and Sustainable Energy Reviews, 2017. **72**: p. 354-362.
86. Kataki, S., et al., *Phosphorus recovery as struvite from farm, municipal and industrial waste: Feedstock suitability, methods and pre-treatments*. Waste Management, 2016. **49**: p. 437-454.

87. Mehta, C.M., et al., *Technologies to Recover Nutrients from Waste Streams: A Critical Review*. Critical Reviews in Environmental Science and Technology, 2015. **45**(4): p. 385-427.
88. Hendriks, A. and J.G. Langeveld, *Rethinking Wastewater Treatment Plant Effluent Standards: Nutrient Reduction or Nutrient Control?* Environ Sci Technol, 2017. **51**(9): p. 4735-4737.
89. Kuenen, J.G., *Anammox bacteria: from discovery to application*. Nat Rev Microbiol, 2008. **6**(4): p. 320-6.
90. Morales, N., et al., *Integration of the Anammox process to the rejection water and main stream lines of WWTPs*. Chemosphere, 2015. **140**: p. 99-105.
91. Wang, S., et al., *Anaerobic ammonium oxidation in traditional municipal wastewater treatment plants with low-strength ammonium loading: Widespread but overlooked*. Water Res, 2015. **84**: p. 66-75.
92. Ma, B., et al., *Biological nitrogen removal from sewage via anammox: Recent advances*. Bioresour Technol, 2016. **200**: p. 981-90.
93. McCarty, P.L., *What is the Best Biological Process for Nitrogen Removal: When and Why?* Environ Sci Technol, 2018. **52**(7): p. 3835-3841.
94. Hu, Z., et al., *Nitrogen removal with the anaerobic ammonium oxidation process*. Biotechnol Lett, 2013. **35**(8): p. 1145-54.
95. Nifong, A., et al. *Performance of a full-scale sidestream demon installation*. in *Proceedings of the WEF Nutrient Removal and Recovery Conference*. 2013. Vancouver, British Columbia, Canada.
96. Lackner, S., et al., *Full-scale partial nitritation/anammox experiences--an application survey*. Water Res, 2014. **55**: p. 292-303.

97. Joss, A., et al., *Combined nitritation-anammox: advances in understanding process stability*. Environmental science & technology, 2011. **45**(22): p. 9735-9742.
98. Jin, R.-C., et al., *The inhibition of the Anammox process: A review*. Chemical Engineering Journal, 2012. **197**: p. 67-79.
99. Tang, C.-J., et al., *Thermodynamic and kinetic investigation of anaerobic bioprocesses on ANAMMOX under high organic conditions*. Chemical engineering journal, 2013. **230**: p. 149-157.

Chapter 2

Pilot Plant Demonstration of Stable and Efficient High Rate Biological Nutrient Removal with Low Dissolved Oxygen Conditions

This chapter is published under the same title:

Keene, N.A., Reusser, S. R., Scarborough, M. J., Grooms, A. L., Seib, M., Santo Domingo, J., Noguera, D. R. *Pilot plant demonstration of stable and efficient high rate biological nutrient removal with low dissolved oxygen conditions*. Water Res, 2017. **121**: p. 72-85. <https://doi.org/10.1016/j.watres.2017.05.029>.

NAK (now NKB), DRN, SRR, ALG, and MS developed the research plan and project goals. NKB started, operated, and collected samples from the pilot-scale bioreactor, conducted batch kinetic tests, performed microscopy, and analyzed sequencing data. NKB and MJS developed the aeration energy model. JSD performed DNA sequencing and contribute to analysis of sequences. All co-authors participated in drafting the manuscript, tables and figures.

Abstract

Aeration in biological nutrient removal (BNR) processes accounts for nearly half of the total electricity costs at many wastewater treatment plants. Even though conventional BNR processes are usually operated to have aerated zones with high dissolved oxygen (DO) concentrations, recent research has shown that nitrification can be maintained using very low-DO concentrations (e.g., below 0.2 mg O₂/L), and therefore, it may be possible to reduce energy use and costs in BNR facilities by decreasing aeration. However, the effect of reduced aeration on enhanced biological phosphorus removal (EBPR) is not understood. In this study, we investigated, at the pilot-scale level, the effect of using minimal aeration on the performance of an EBPR process. Over a 16-month operational period, we performed stepwise decreases in aeration, reaching an average DO concentration of 0.33 mg O₂/L with stable operation and nearly 90% phosphorus removal. Under these low-DO conditions, nitrification efficiency was maintained, and nearly 70% of the nitrogen was denitrified, without the need for internal recycling of high nitrate aeration basin effluent to the anoxic zone. At the lowest DO conditions used, we estimate a 25% reduction in energy use for aeration compared to conventional BNR operation. Our improved understanding of the efficiency of low-DO BNR contributes to the global goal of reducing energy consumption during wastewater treatment operations.

Introduction

Traditional high-rate biological nutrient removal (BNR) processes operate with extensive aeration to ensure efficient nitrification and phosphorus (P) removal. Since aeration is one of the most energy intensive activities at BNR facilities, accounting for 30 to 60% of total energy consumption [1], a reduction in aeration intensity without negatively impacting nutrient removal is desirable.

It is well established that simultaneous nitrification and denitrification (SND) occurs in extended aeration processes in zones having low dissolved oxygen (DO) concentrations [2]. For instance, oxidation ditch-type reactors have been shown to achieve up to 50% total nitrogen removal in sections of the process with low aeration [3, 4]. Membrane bioreactors have also been efficiently operated with low-DO to achieve SND [5, 6]. More recent operational approaches such as ammonia-based aeration control [7-9] have also been shown to effectively achieve SND at low-DO.

Recent experiments have demonstrated that achieving efficient low-DO nitrification requires adaptation of the microbial communities in the activated sludge. Although nitrification ceased when samples from a conventional activated sludge plant were transferred to low-DO conditions (0.2 mg O₂/L), nitrification capacity of the sludge was restored after prolonged exposure of the culture to this low-DO environment [10]. When DO was slowly reduced in a bench-scale system, nitrification rates were not impacted; however, the adaptation of nitrifying organisms to low-DO conditions was dependent on the sludge age of the system [11].

Although nitrification at low-DO has been well documented, from lab-scale to full-scale, studies reporting long-term enhanced biological phosphorus removal (EBPR) at low-DO are infrequent. Efficient operation of lab-scale EBPR systems with DO in the aerated zone maintained below 0.45 mg O₂/L has been reported [12-14]. In addition, kinetic batch experiments suggest that polyphosphate accumulating organisms (PAO) have a higher affinity for oxygen than glycogen accumulating organisms (GAO), and therefore, that low-DO operation may favor PAO activity [15].

In this project we studied the effect of steady reductions in aeration in a high-rate BNR pilot plant that simulated the University of Cape Town (UCT)-type configuration used at the Nine Springs Wastewater treatment plant (WWTP) (Madison, WI). The pilot plant, fed primary effluent from the full-scale facility and operated for a period of 16 months, initially simulated high aeration conditions, and then underwent stepwise decreases in aeration. We show that it was possible to achieve efficient nitrification and EBPR with average DO of 0.33 mg O₂/L, and that the process maintained efficient nutrient removal when aeration was decreased in a stepwise manner. Batch kinetic experiments of nitrification and P uptake at a range of DO concentrations were performed to evaluate the adaptation of the microbial community to minimal aeration. In addition, we assessed the succession in nitrifying and P removing organisms as a result of decreasing DO with 16S rRNA gene sequencing analysis.

Materials and Methods

Continuous flow pilot-scale reactor description and operation

A continuous flow pilot-scale treatment train was seeded from the Nine Springs WWTP (Madison, WI) and was configured to mimic the full-scale process (Figure 2-1), which operates as a modified UCT without internal nitrate recycle (since this plant is not required to achieve denitrification). The full-scale Nine Springs WWTP operates with a solids retention time (SRT) and a hydraulic retention time (HRT) of approximately 10 days and 17 hours, respectively. The DO in the full-scale aerobic zone reaches concentrations greater than 2.0 mg O₂/L. The operational parameters for the pilot- and full-scale plants are summarized in Table S2-1.

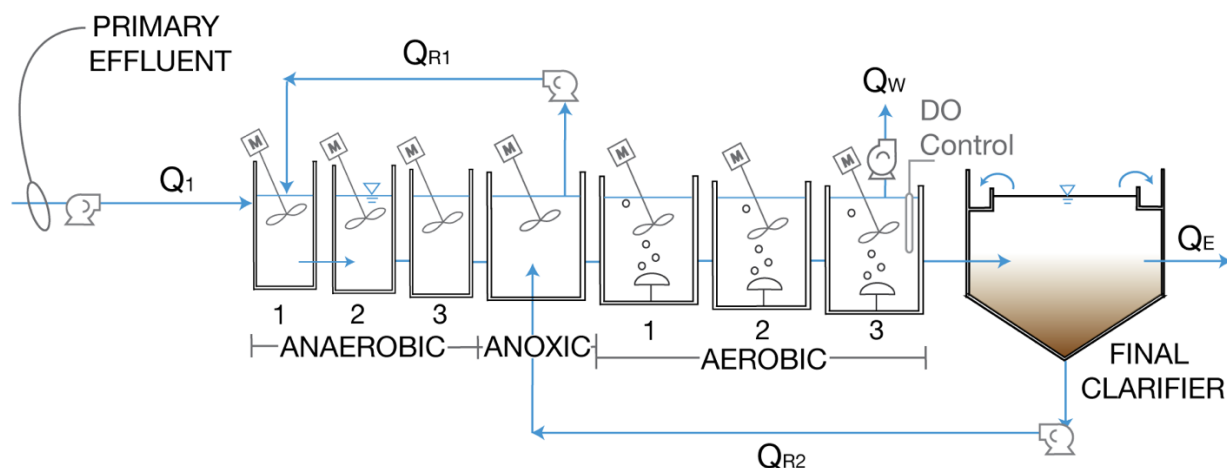


Figure 2-1. Continuous flow pilot-scale treatment train operated as a modified UCT process without nitrate recycling, simulating the configuration of the full-scale Nine Springs WWTP. The anaerobic (3 tanks), anoxic (1 tank), and aerobic (3 tanks) portions of the system were equipped with mechanical mixers. The internal recycle (QR1) returned biomass to the beginning of the system from the anoxic tank. Return activated sludge (QR2) was recycled from the bottom of the clarifier to the anoxic zone. Wasting (QW) occurred from the last aerobic tank.

The anaerobic, anoxic, and aerobic portions of the pilot reactor consisted of seven tanks with a combined total volume of 495 gallons (1870 L). Each tank was equipped with a mechanical mixer. In addition, the system included a secondary clarifier with a total volume of 268 gallons (1010 L). Primary effluent from the full-scale WWTP was directly used as influent to the pilot-scale treatment train. The influent flow rate varied in the range of 1200 to 1340 gallons per day (GPD) ($4.54\text{--}5.08\text{ m}^3/\text{d}$) and produced a hydraulic retention time of approximately 9 to 10 hours. The anaerobic recycle was maintained at an equivalent flow rate to the influent flow, simulating the operational condition at the full-scale plant. Return activated sludge (RAS) was pumped to the anoxic zone at an average of 2190 GPD ($8.28\text{ m}^3/\text{d}$). The solids retention time (SRT) was maintained at an average of 10 ± 3.4 days by wasting mixed liquor from the third aerobic tank.

DO was measured in real-time using an optical probe connected to a DO meter (YSI IQ SensorNet FDO Optical DO sensor and YSI IQ SensorNet 182 meter, Yellow Springs, OH) and interfaced to a desktop computer. Data acquisition was performed using LABVIEW (National Instruments, Austin, TX) with DO concentrations recorded every 20 minutes. Air was provided, from a tap off of the air supply to the full-scale treatment plant, through a fine-bubble membrane diffuser (Sanitaire Silver Series II, Brown Deer, WI) installed at the base of each aerobic tank. Airflow was controlled by direct-reading variable area flow meters (King Instrument Co, Garden Grove, CA) and set at approximately 1.5 SCFM (43 L/min), 1.0 SCFM (28 L/min), and 0.5 SCFM (14 L/min) for the first, second, and third aerobic tanks, respectively. The DO meter maintained a

pre-selected maximum DO concentration set point in one of the tanks, and air supply to all of the aerated tanks was controlled via a single solenoid valve.

The reactor was operated in three different phases. In the first phase (days 0-181), DO was maintained relatively high (i.e., average of 0.51 mg O₂/L in the first tank, 0.72 mg O₂/L in the second tank, and 1.11 mg O₂/L in the third tank). Subsequently, DO concentrations in the aerated tanks were steadily decreased in the second operational phase (days 182-365), with at least 3 weeks of operation in between DO reductions. During the third operational phase (days 366-493), the reactor was allowed to stabilize at the lowest DO conditions achieved (i.e., averages of 0.23, 0.48, and 0.28 mg O₂/L in the first, second, and third tank, respectively).

Sample preparation and analysis

Grab samples of primary effluent and pilot plant effluent were collected approximately 2 to 3 times per week. Grab samples from each tank in the pilot plant were collected weekly. Approximately 15 mL of each sample was immediately filtered through a 0.45- μ m membrane filter (Nitrocellulose Membrane Filters, EMD Millipore Corp., Darmstadt, Germany) and the filtrate stored at 4 °C until further analysis. The remaining unfiltered sample was used for solids measurement and DNA extraction.

Total suspended solids (TSS), volatile suspended solids (VSS), phosphate (PO₄³⁻), total ammoniacal nitrogen (TAN; NH₄⁺-N plus NH₃-N), Total Kjeldahl Nitrogen (TKN) were conducted following standard methods [16]. Nitrite (NO₂⁻-N) and nitrate (NO₃⁻-N) were measured with high performance liquid chromatography (HPLC) using a Restek Ultra

Aqueous C18 column (Restek Corporation, Bellefonte, PA) and detection by UV at 214 nm in a Shimadzu HPLC system (Shimadzu Scientific Instruments, Columbia, MD).

Kinetic tests

Nitrification, denitrification, and P uptake experiments were performed in 3-L batch reactors using sludge from the pilot- and full-scale reactors. Activated sludge was collected from the anaerobic zone, stored anaerobically for 20-30 minutes during transport to the laboratory, and subsequently equilibrated to room temperature.

During ammonia oxidation and P uptake experiments, air was pumped into the reactors through an air bubble stone with DO monitored and controlled every 0.1 minutes. Targeted DO concentrations were between 0.1 and 4.0 mg O₂/L. Measured substrate utilization rates during the batch experiments were normalized by the VSS concentrations. The effect of DO concentration was simulated using a Monod kinetic model [17]. Air was not supplied during denitrification experiments. Instead, either sodium nitrite or sodium nitrate were added to create initial nitrite and nitrate concentrations of approximately 20 mg N/L.

Floc and microcolony size measurements

Full-scale plant and pilot plant floc and microcolony sizes were measured and compared to determine impacts from low-DO operation. A detailed summary of this analysis is described in section 3.1 of the supplementary document.

16S ribosomal RNA gene tag sequencing

High-throughput sequencing of 16S rRNA gene fragments was used to analyze the microbial communities in samples from the pilot- and full-scale plants. Links to the raw sequence data and sample metadata described in this study can be accessed through the NCBI BioProject database using the accession PRJNA358610. During the 16-months of pilot plant operation, a total of 53 and 51 grab samples were collected from the pilot- and full-scale plants, respectively. Biomass sample collection began on day 120 of pilot plant operation and continued on a weekly basis through day 490. DNA was extracted using Power® Soil DNA Isolation Kit (MoBIO Laboratories, Carlsbad, CA). Extracted DNA was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and stored at -20 °C. Barcoded PCR primers F515/R806 [18] were used to amplify the V4 hypervariable region of the 16S rRNA gene. Purified amplicons were pooled in equimolar quantities and sequenced on an Illumina MiSeq sequencer (Illumina, San Diego, CA), using pair-end 250 base pair kits, at the Cincinnati Children's Hospital DNA core facility.

Files containing paired-end reads were quality trimmed and filtered with Sickle Paired End ('sickle pe') using the Sanger fastq file quality type and default minimum length and quality thresholds [19]. The quality trimmed paired-end reads were merged with Fast Length Adjustment of Short Reads (FLASH) version 1.2.11 using default parameters [20]. Then, the merged reads were aligned, filtered, and binned into operational taxonomic units (OTU) with 97% identity using the standard QIIME protocol [21]. USEARCH version 6.1 was used for OTU picking and chimera detection [22]. The

representative sequences from each OTU were taxonomically classified using the MIDas-DK database [23].

To identify and compare core taxa, OTUs were summarized and normalized at the genus level using a QIIME workflow script (i.e., 'summarize_taxa_through_plots.py'). To analyze the dataset further, the OTU table along with associated metadata and representative sequences, were imported, stored, and subset using version 1.14.0 of the Phyloseq package [24] and the R statistical framework, version 3.2.4 [25]. Comparisons of relative abundance between samples during ordination and dendrogram analysis were made after the dataset was rarefied to an even depth (9,693 reads per sample) and filtered to remove low abundance OTUs (i.e., less than 3 sequences). The Bray-Curtis method [26] was used to compute the dissimilarity matrices for downstream principle coordinate analysis (PCoA). PCoA ordination was calculated and plotted onto two axes using Phyloseq functions (e.g., 'ordinate' and 'plot_ordination'). Analysis of similarities (ANOSIM), with 999 permutations, was used to statistically test whether there were significant differences between groups (full-scale, and pilot plant operational phases 1, 2, and 3) shown in the PCoA ordination [27]. Graphics were generated with the ggplot2 [28] and gplots R packages [29]. Representative sequences were further classified using the Basic Local Alignment Search Tool (BLAST) [30]. Multiple sequence alignments were performed using the Geneious global alignment with free end gaps in Geneious version 9.1.2 [31]. Consensus phylogenetic trees were produced using the Neighbor-Joining method on a Tamura-Nei genetic distance model with a bootstrap resampling method in Geneious [31].

Energy Savings Calculations

Aeration energy requirements were determined for the full-scale and pilot-scale systems using existing data from the Nine Springs WWTP. A detailed summary of this analysis is described in section 3.2 of the supplementary document.

Results

Reactor performance during step reductions in aeration

Over all three operational phases, the influent phosphate (P) concentration was 6.5 ± 1.9 mg PO_4^{3-} -P/L. Influent TKN and TAN was 38 ± 5.7 mg TKN/L and 32 ± 5.7 mg NH_3 -N/L, respectively. During the first operational phase (day 0 to 181, Figure 2-2), average DO concentrations in the three aerated tanks were 0.51 ± 0.28 , 0.72 ± 0.44 , and 1.11 ± 0.28 mg O_2 /L, respectively. Effluent P was 0.55 ± 0.33 mg PO_4^{3-} -P/L toward the end of this operational period, resulting in an average P removal efficiency of 90 ± 7 %. Effluent TKN and TAN concentrations were 1.3 ± 0.51 mg TKN/L and 0.75 ± 0.54 mg NH_3 -N/L, resulting in removal efficiencies of 96% and 98%, respectively.

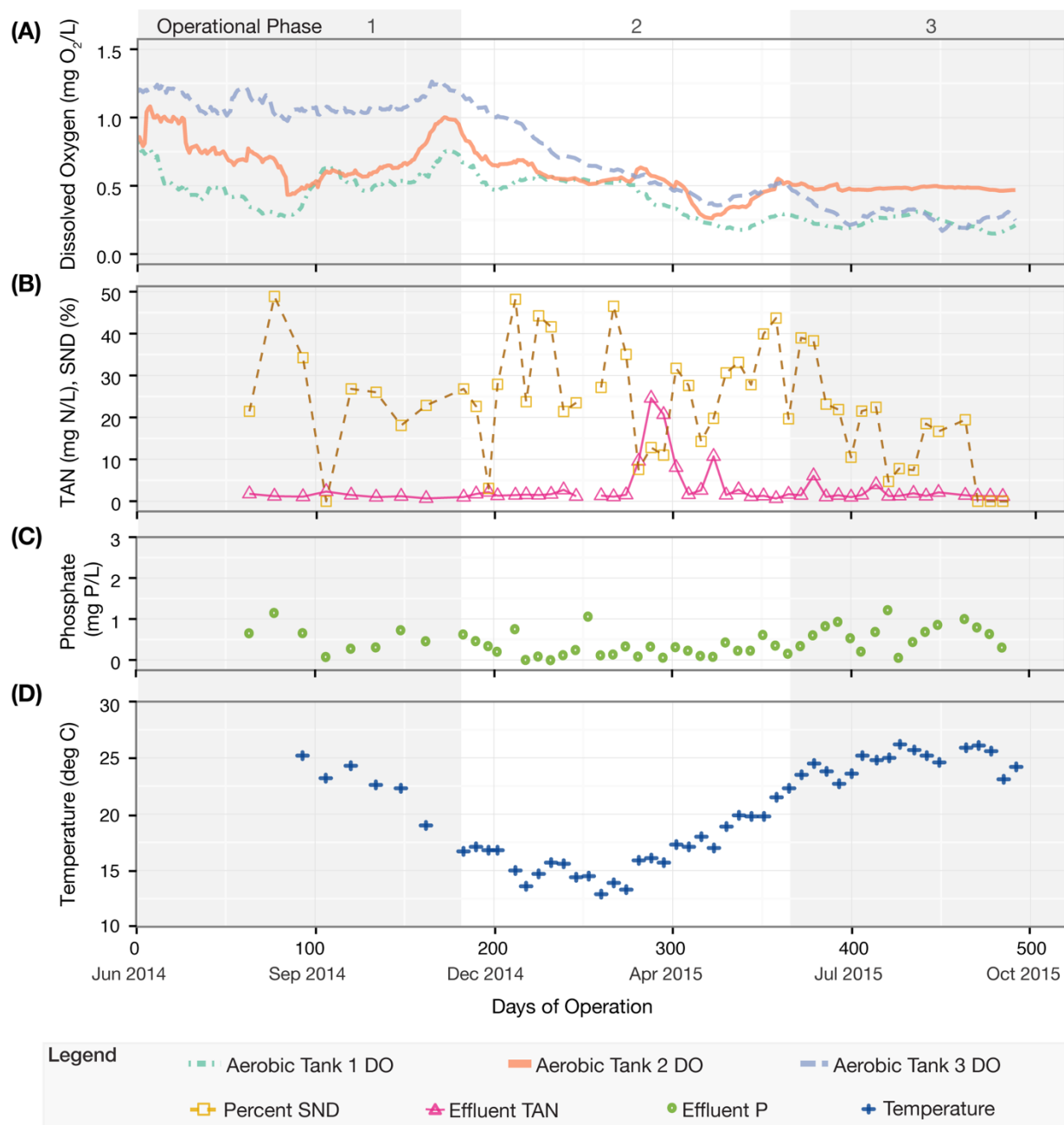


Figure 2-2. Pilot plant performance during the three operational phases. (A) Dissolved oxygen (DO), represented as a 20-point moving average time series from the three aerobic tanks. (B) Effluent Total Ammoniacal Nitrogen (TAN) and percent of simultaneous nitrification and denitrification (SND). (C) Effluent phosphate (P). (D) Activated sludge temperature.

The second operational phase (days 182 to 365, Figure 2-2) was characterized by a steady decrease in the DO concentrations in the three aerated tanks. P removal remained stable during this period of operation, with effluent concentrations of 0.30 ± 0.24 mg PO_4^{3-} -P/L, corresponding to a 95 ± 3 % P removal efficiency. During this phase, nitrogen removal was sensitive to mechanical disruptions in plant operation. For instance, at day 284, the pilot plant experienced an overflow event that resulted in significant loss of biomass and a decrease in the nitrification efficiency for approximately 18 days (Figure 2-2B). Less severe mechanical disruptions occurred at days 278 and 314, which were also followed by temporary decreases in nitrification efficiency. To accelerate the recovery of nitrification in the reactor after these operational upsets, biomass wasting was temporarily discontinued. Excluding the days where nitrification was impacted by operational failure, the TKN and TAN effluent concentrations were an average of 1.5 ± 0.52 mg TKN/L and 0.55 ± 0.54 mg NH_3 -N/L over this intermediate period of operation.

During the final 127 days of operation (day 366-493), a constant DO concentration was maintained in each of the three aerated tanks (average 0.23 ± 0.13 , 0.48 ± 0.12 , and 0.28 ± 0.25 mg O_2 /L, respectively). P removal efficiencies were at 89 ± 9 %, with effluent P concentrations of 0.77 ± 0.62 mg PO_4^{3-} -P/L. TKN and TAN removal efficiencies were greater than 96%, with effluent concentrations at an average of 1.4 ± 0.30 mg TKN/L (excluding two days impacted by minor mechanical disruptions, at days 379 and 414) and 0.13 ± 0.11 mg NH_3 -N/L, respectively.

Extent of simultaneous nitrification and denitrification (SND) during reactor operation

The total nitrogen removal efficiency was high during the operation of the reactor ($70 \pm 10\%$). While influent TKN concentrations were 38 ± 5.7 mg TKN/L over all three phases of operation, effluent nitrate plus nitrite concentrations (nitrite below detection limit most of the time) were 8.6 ± 1.7 mg N/L, 8.0 ± 1.9 mg N/L, and 10 ± 1.4 mg N/L for each of the three operational phases, respectively. This level of nitrogen removal was consistently higher than observed in the full-scale plant during the same period of operation, which had an average monthly effluent nitrate plus nitrite concentration of 20 ± 1.4 mg N/L. The difference could be due to denitrification improvements in either the anoxic tank (Figure 2-1) or in the low-DO tanks (via SND). Using the difference of TKN in the anoxic stage and total nitrogen in the effluent as a proxy for SND in the plant, we estimated the extent of SND throughout the study (Figure 2-2B). This metric fluctuated widely, with sampling days in which minimal SND was observed (e.g., days 197 and 471 - 492) and days when SND was higher than 40% (e.g., days 212, 267 and 358). During the last operational phase, when DO was the lowest, the variability decreased, and SND was markedly lower than in the earlier two phases. Nonetheless, the total nitrogen removal, which includes SND plus denitrification in the anoxic tank, remained high ($68 \pm 6\%$) during the third operational phase.

In order to evaluate the potential contribution of denitrifying polyphosphate accumulating organisms (DPAO) to SND in the reactor, we performed batch tests to detect P uptake in the presence of nitrate or nitrite, and in the absence of oxygen (Figure 2-3).

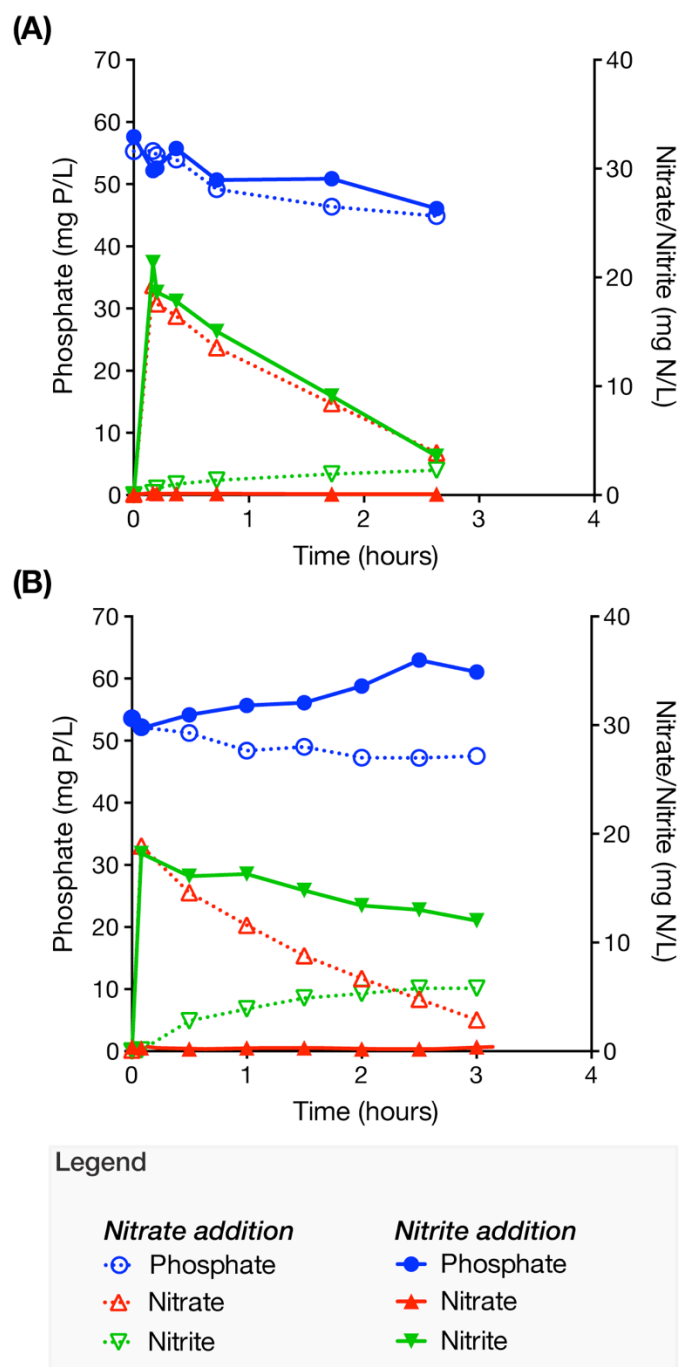


Figure 2-3. Batch experiments of anoxic phosphorus uptake in the presence of nitrate or nitrite, (A) during high-DO operation, and (B) after stable low-DO operation.

Figure 2-3A shows that at the end of the first operational phase denitrification was possible using nitrate or nitrite as electron acceptors. Approximately 8.8 mg N/L of nitrate and 14 mg N/L of nitrite were denitrified with simultaneous P uptake, at rates of 1.4 and 0.81 mg P h⁻¹ gVSS⁻¹, respectively. After stable low-DO operation (Figure 2-3B), a similar extent of denitrification was observed when nitrate was the electron acceptor (i.e., 7.3 mg N/L denitrified), with a P uptake rate of 0.99 mg P h⁻¹ gVSS⁻¹. In contrast, denitrification decreased by a factor of 2.4 when nitrite was the electron acceptor (i.e., 5.9 mg N/L denitrified) and no P uptake was observed. For comparison, the measured P uptake rate in the pilot plant during stable low-DO operation was 5.0 ± 1.9 mg P h⁻¹ gVSS⁻¹, which is approximately five times greater than observed in the nitrate test. Thus, the results suggest that the majority of the P removal in the pilot plant was not associated with DPAO, and therefore, that DPAO were not significant contributors to SND in the reactor.

Microbial adaptation to low-DO

The effect of DO on nitrification and P uptake was studied using batch experiments. For comparison, ammonia oxidation and P uptake experiments were performed with samples from the full-scale Nine Springs WWTP and the pilot plant during the third operational phase, when DO was at the lowest levels used.

The maximum specific ammonia oxidation rates were 0.11 and 0.10 mg N/mg VSS-d for the full-scale and pilot-scale samples, respectively, whereas the half saturation constants were 1.38 and 0.30 mg O₂/L for the full-scale and the pilot plant (Figure 2-4A; Table 2-1). Thus, the adaptation of the ammonia oxidizing microbial community to

low-DO in the pilot plant was reflected in the observed half saturation constant, which was an order of magnitude lower at low-DO conditions compared to full-scale operation. Additionally, in these experiments, ammonia was completely oxidized to nitrate, demonstrating that the nitrite oxidizing community was also adapted to the low-DO conditions. This is consistent with the pilot plant operation, where nitrite was rarely detected, and nitrate was continuously produced throughout the three phases of operation.

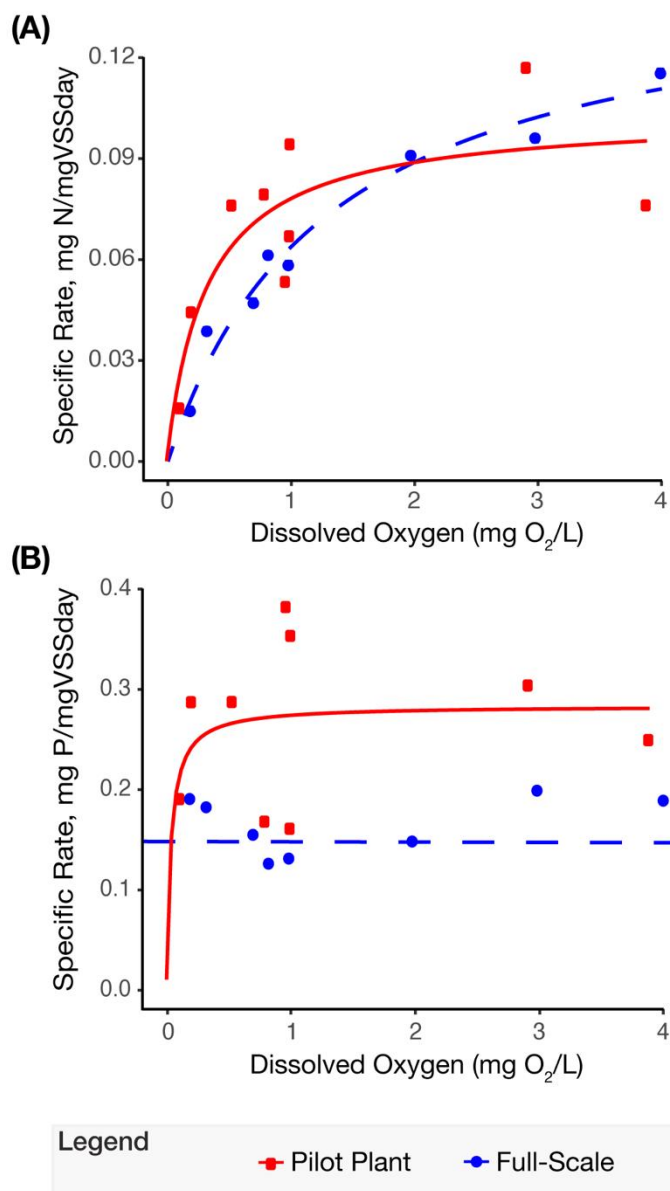


Figure 2-4. (A) Comparison of specific rate of ammonia oxidation of activated sludge from the full-scale plant, and from the pilot-scale plant, after stable low-DO operation. Points represent experimental measurements of specific rates at different DO concentrations. Lines represent the best-fit of a Monod kinetic model to the experimental data. (B) Specific rates of phosphorus uptake of activated sludge from the full-scale plant and from the pilot-scale plant after stable low-DO operation (phase 3). Symbols represent experimental measurements of specific rates at different DO concentrations. A Monod kinetic model (solid line) could only be fitted to the data from the pilot-scale plant. A dashed line represents the average maximum specific rate for the full-scale plant.

Table 2-1. Best-fit Monod kinetic parameters for ammonia oxidation and phosphorus uptake from the full-scale plant and from the pilot-scale plant after stable low-DO operation.

		Sample Analyzed		
		Units	Full-Scale	Pilot Plant Operational Phase 3
Average VSS		mg/L	930 ± 270	1640 ± 600
Ammonia Oxidation	$q_{\max}$	mg/mgVSSday	0.11 ± 0.010	0.10 ± 0.013
	K_{DO}	mg O ₂ /L	1.38 ± 0.23	0.30 ± 0.14
	R^2	none	0.98	0.80
Phosphorus Uptake	$q_{\max}$	mg/mgVSSday	0.16 ± 0.018	0.36 ± 0.058
	K_{DO}	mg O ₂ /L	-	0.091 ± 0.085
	R^2	none	-	0.32

A similar kinetic analysis was performed to evaluate the effect of DO on P uptake under various DO concentrations (Figure 2-4B). Interestingly, the specific rate of P uptake in samples from the full-scale and pilot-scale plants did not change as a function of DO. A half saturation constant could only be determined for the pilot plant ($K_{DO} = 0.09$ mg O₂/L), whereas the Monod model was not adequate to represent the effect of oxygen on P uptake in samples from the full-scale plant. However, the maximum specific P uptake rate in samples from the pilot plant was roughly twice the uptake rate that of the full-scale samples (Table 2-1). Thus, polyphosphate accumulating organisms (PAO) had a high affinity for oxygen in both plants, and adaptation to low-DO was reflected in higher uptake rates and not in higher oxygen affinity.

In activated sludge, floc and microcolony sizes can impact a microorganisms apparent affinity for substrates, such as oxygen, due to diffusion gradients [32]. In order to determine whether the observed oxygen affinity could have been influenced by changes

in mass transfer limitations, we measured the size distribution of activated sludge flocs and of ammonia oxidizing bacteria (AOB) microcolonies. After reducing DO, the median floc and AOB microcolony sizes decreased and overall size distribution narrowed (Figure 2-5). Median floc equivalent diameter in the full-scale and pilot-scale plant were 152 μm (interquartile range, IQR = 113 - 192 μm) and 117 μm (IQR = 87 - 151 μm), respectively (Figure 2-5A). Median AOB microcolony equivalent diameter in the full-scale and pilot-scale plant were 8.9 μm (IQR = 6.5 - 11.4 μm) and 5.6 μm (IQR = 4.4 - 7.1 μm), respectively (Figure 2-5B). While median floc size decreased by only 20% after low-DO operation, this decrease was statistically significant ($p < 0.0001$ two tailed, Mann-Whitney $U=43309$) [33]. Additionally, median AOB microcolony size decreased by nearly 40% in the low-DO pilot plant ($p < 0.0001$ two-tailed, Mann-Whitney $U=2464$). Therefore, the observed higher affinity for oxygen in the pilot-scale plant may be partially explained by a reduction in mass transfer limitations within the flocs and the AOB microcolonies.

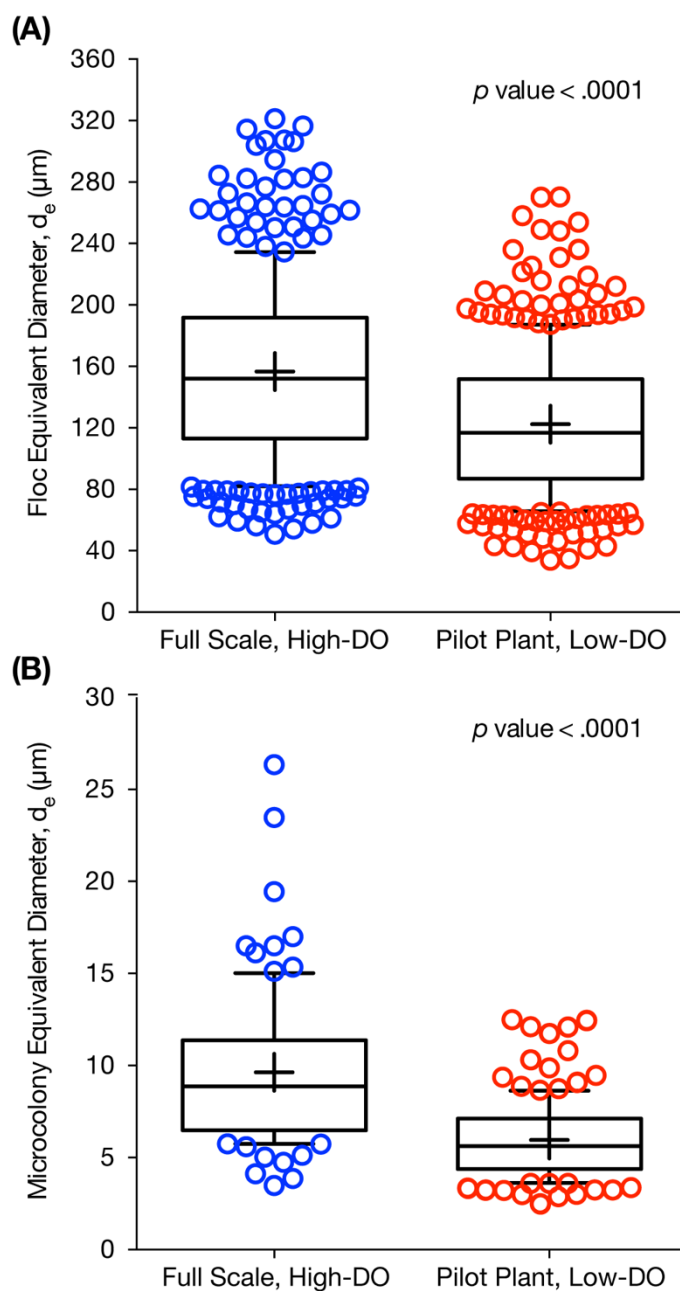


Figure 2-5. Box plots comparing (A) floc size distribution (by median), and (B) AOB microcolony size distribution (by median) for the full-scale plant operating with high DO concentrations and the pilot plant during low-DO operation. The upper and lower bounds of the boxes represent the 25th and 75th percentiles, the whiskers denote the 10th and 90th percentiles, the circular symbols are the outliers, and the plus symbol represents the mean.

Microbial community composition and dynamics

Next-generation 16S rRNA sequencing was performed with samples from the pilot and full-scale plants to gain further insight into the microbial populations responsible for nitrification and EBPR under low-DO conditions. A total of 2.7 million reads from 53 pilot plant samples and 2.2 million reads from 51 full-scale samples passed all quality checks and were distributed across 76,452 OTUs defined at 97% sequence identity. The core taxa (i.e., populations that were always present) in both plants were identified by initially binning OTUs into groups based on identical taxonomic assignments, which resulted in a total of 53 phyla, 138 classes, and 1,383 genera representing all OTUs. Using a mean relative abundance greater than 0.5% as the cut off to define abundant core taxa, 45 genera from the full-scale plant, representing 65% of the total reads, were considered the abundant high-DO core taxa, while 37 genera were categorized as the abundant low-DO core taxa, and accounted for 69% of the total reads from the pilot plant during the final phase of operation. Twenty-five genera were core taxa common to both plants. The 30 most abundant core taxa for the pilot and full-scale plants are shown in Figure 2-6.

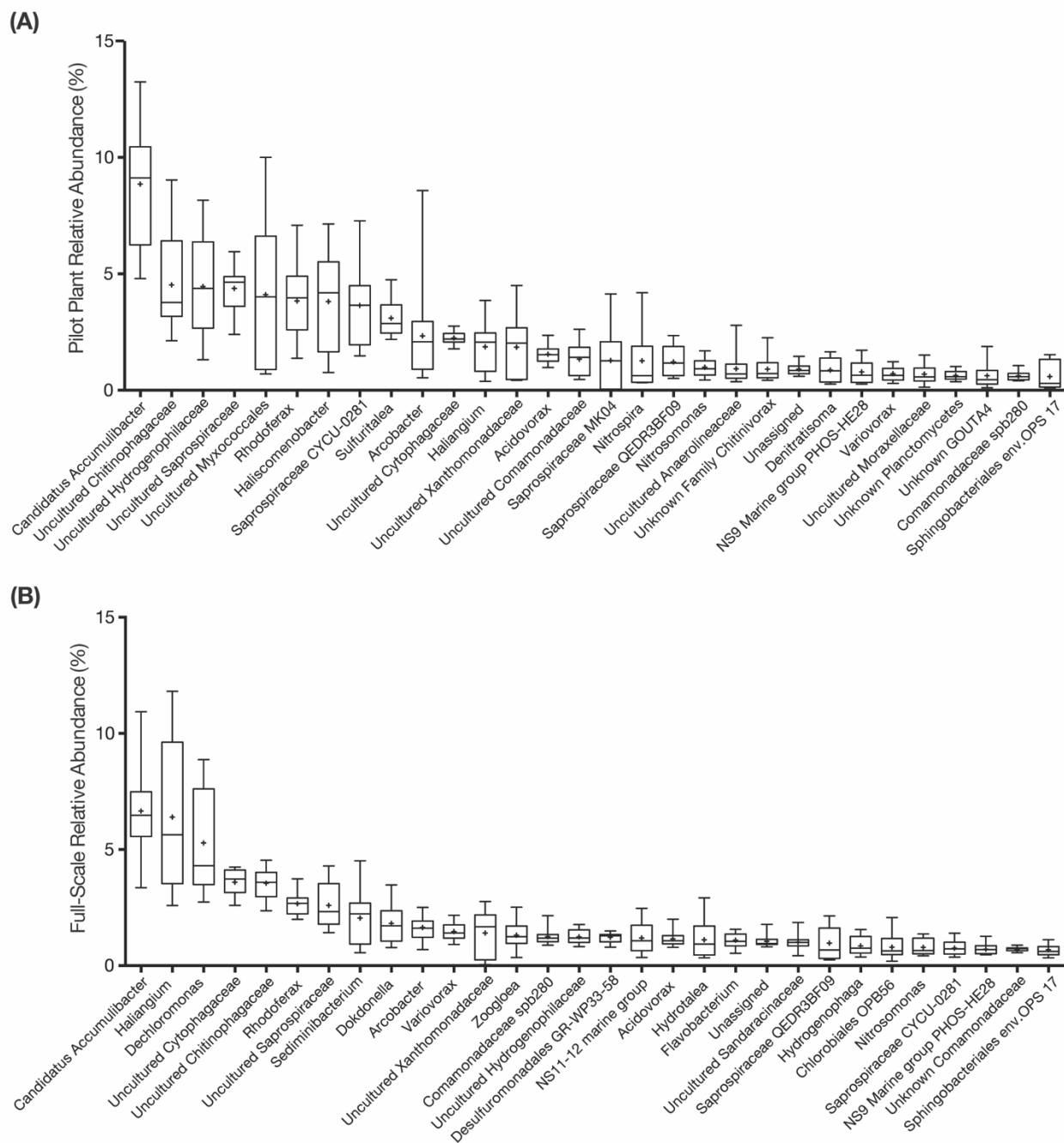


Figure 2-6. Box plots of the 30 most abundant core taxa classified at the genus level (by median) for (A) Pilot plant, and (B) Full-scale plant during the low-DO operational phase (phase 3). The names in the horizontal axis are the lowest assigned taxonomic rank. The upper and lower bounds of the boxes represent the 25th and 75th percentiles, the whiskers denote the maximum and minimum values, and the plus symbol represents the mean.

Candidatus Accumulibacter phosphatis (*Ca. Accumulibacter*) was the most abundant organism detected in both plants (Figure 2-6), with mean relative abundances of $5.9 \pm 3.5\%$ and $4.5 \pm 2.4\%$ in the pilot- and full-scale plants, respectively. At the genus-level, although there was seasonal variation of *Ca. Accumulibacter*, with lower abundance in winter and spring (Figure 2-7A), the low-DO of the pilot plant did not negatively impact its relative abundance ($8.9 \pm 2.6\%$ and $6.7 \pm 1.9\%$ in pilot and full-scale). Given its high relative abundance, this PAO was likely responsible for most of the P removal in both plants. The genus *Tetrasphaera*, recently described as containing PAO contributing to P removal in WWTPs [34], was detected but with mean relative abundances of $0.48 \pm 0.46\%$ in full-scale and $0.14 \pm 0.12\%$ in the pilot plant (Figure 2-7B), below the cut off criteria for core taxa and at least an order of magnitude lower than *Ca. Accumulibacter*. Moreover, GAO were also detected, but also at low abundance. *Candidatus Competibacter*, *Defluviicoccus*, and *Propionivibrio* were present at mean relative abundances less than 0.2%.

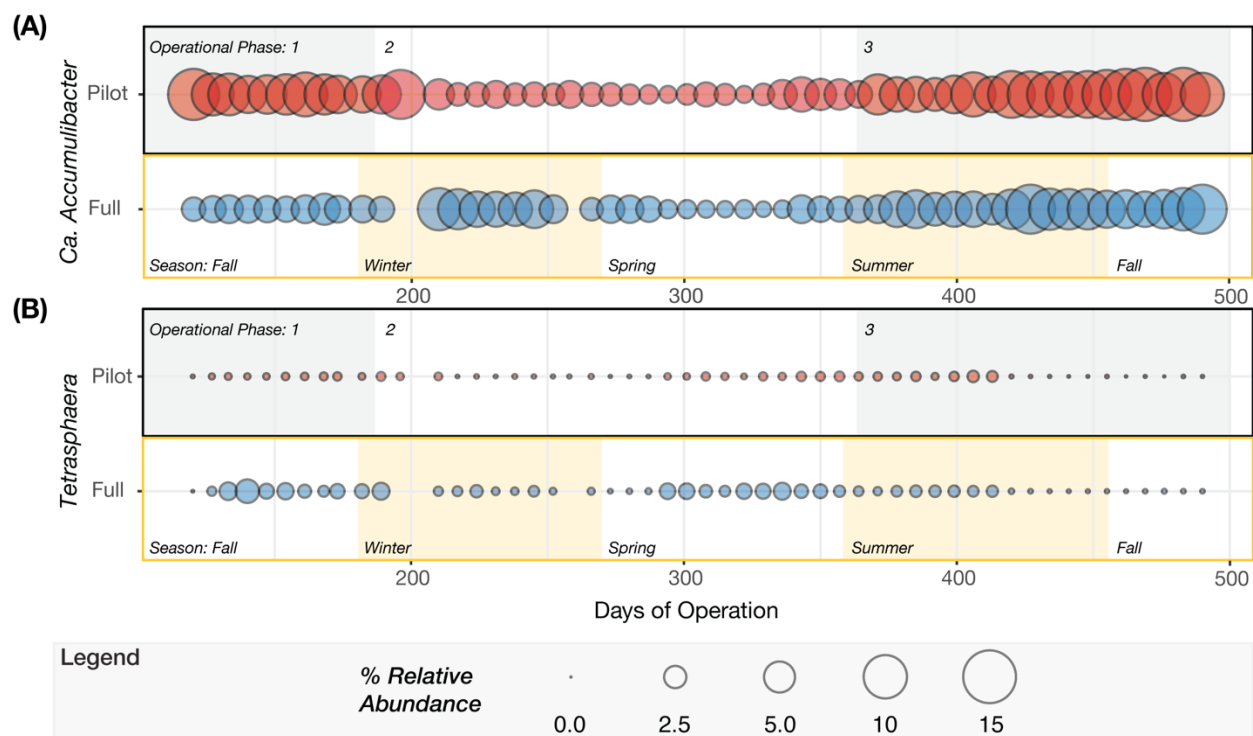


Figure 2-7. Balloon plots representing temporal variation in relative abundance of genera known to perform phosphorus cycling in EBPR reactors, **(A)** *Candidatus Accumulibacter* and **(B)** *Tetrasphaera*. For each genus, the pilot plant is described in the upper panel and full-scale plant in the lower panel. Relative abundance was determined based on the entire community. The shaded areas in the upper panels denote the three operational phases, whereas the shaded areas in the lower panels indicate the meteorological seasons between days 120 and 490 of pilot plant operation.

To further compare and evaluate the relationship of *Ca. Accumulibacter* among samples from both plants, PCoA ordination was performed (Figure 2-8). When all OTUs are included, and samples are categorized into four groups (full-scale, and pilot plant operational phase 1, 2 and 3), the analysis indicates that the overall microbial communities were different (ANOSIM $P = 0.001$, $R = 0.83$, $n_{\text{otu}} = 8,944$) (Figure 2-8A). Here, the R test statistic >0.75 indicates that the groups are “well separated” [35]. In contrast, when only OTUs classified as *Ca. Accumulibacter* are used in the analysis, a scattered distribution is obtained, without major differentiation between sample groups (ANOSIM $P = 0.002$, $R = 0.13$, $n_{\text{otu}} = 603$) (Figure 2-8B). Although the p -value suggests that the differences are statistically significant, the R test statistic <0.25 corresponds to “barely separable” groups [35]. Further analysis of OTUs classified as *Ca. Accumulibacter* revealed that only two of these OTUs were consistently present in the two plants over the entire period of the study (Figure S2-1). The representative sequences of these two prevalent *Ca. Accumulibacter* OTUs (i.e., denovo7 and denovo50 in Figure S2-1) had 99% and 97% sequence identity to *Ca. Accumulibacter* Clade IIA str. UW-1 [36], but the 16S rRNA sequence information was not sufficient to further classify the OTUs into the polyphosphate kinase-based clades that are traditionally used for *Ca. Accumulibacter* classification [37].

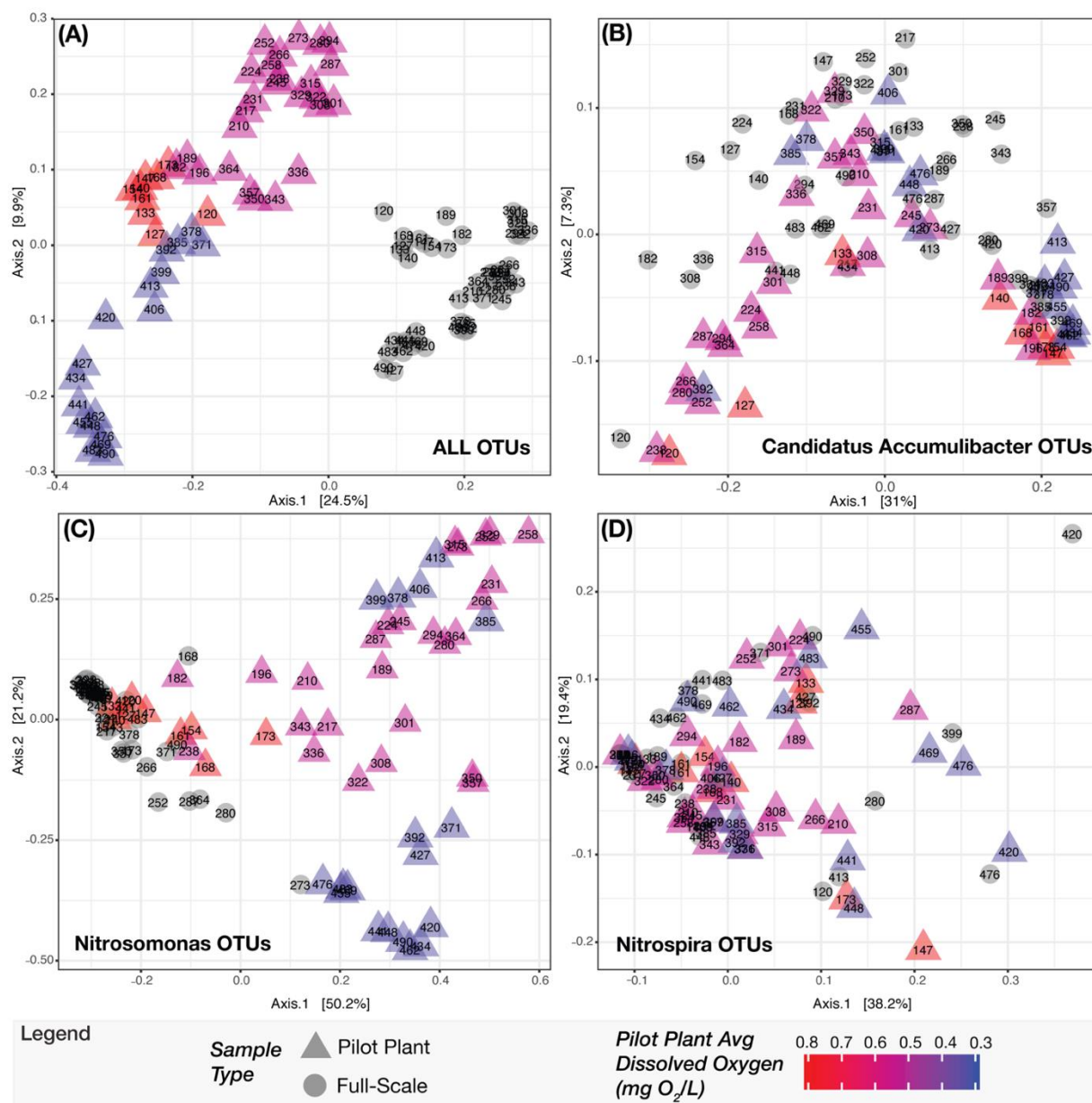


Figure 2-8. Principle coordinate analysis (PCoA) ordination of next-generation sequence data comparing sample composition between full-scale and pilot-scale plants using Bray-Curtis dissimilarity matrices. Full-scale samples and pilot plant samples are represented as circles and triangles, respectively. The sample label represents the day of pilot plant operation that the sample was collected. The color gradient denotes the average DO for the pilot plant samples during each operational phase. (A) PCoA with all OTUs from the rarefied and filtered dataset considered. (B) PCoA with OTUs classified as the genera *Candidatus Accumulibacter*, and with relative abundance calculated using only the total number of reads in the *Ca. Accumulibacter* specific OTUs. (C) PCoA with relative abundance calculated using only the *Nitrosomonas* specific OTUs. (D) PCoA with relative abundance calculated using only the *Nitrospira* specific OTUs.

The genus *Nitrosomonas* was the most abundant AOB detected in both plants (Figure 2-6). During the third phase of operation, the mean relative abundance of *Nitrosomonas* in the pilot- and full-scale plants was $1.0 \pm 0.36\%$ and $0.79 \pm 0.33\%$, respectively. Other AOB, such as *Nitrosomonadaceae*, *Nitrospira* and *Nitrosococcus*, as well as ammonia oxidizing archaea (AOA) were infrequently detected with very low relative abundance (*Nitrosomonadaceae* less than 0.03%, *Nitrospira* and *Nitrosococcus* less than 0.001%, AOA less than 0.0005%) in both plants. While temporal variations of *Nitrosomonas* in the full- and pilot-scale plants were not distinguishable at the genus level (Figure 2-9A), PCoA ordination revealed that the *Nitrosomonas* community in the pilot plant was similar to full-scale at the beginning of the pilot plant operation, but diverged as the DO decreased (ANOSIM $P = 0.001$, $R = 0.73$, $n_{\text{otu}} = 45$) (Figure 2-8C). Here, the R test statistic > 0.5 indicates that the groups are “separated but overlapping” [35]. Further analysis of OTUs classified as *Nitrosomonas* revealed one OTU (i.e., denovo56 in Figure S2-2) was dominant in the full-scale plant over the entire time series and in the pilot plant during the first operational phase but decreased in abundance in the pilot plant as the DO was decreased (Figure S2-2). At the lower DO, other *Nitrosomonas* OTUs (i.e., denovo201 and denovo166) became enriched (Figure S2-2) in the pilot plant. A phylogenetic analysis based on the 16S rRNA gene sequences (Figure 2-10) indicated that denovo56 shared a 99.3% sequence identity with

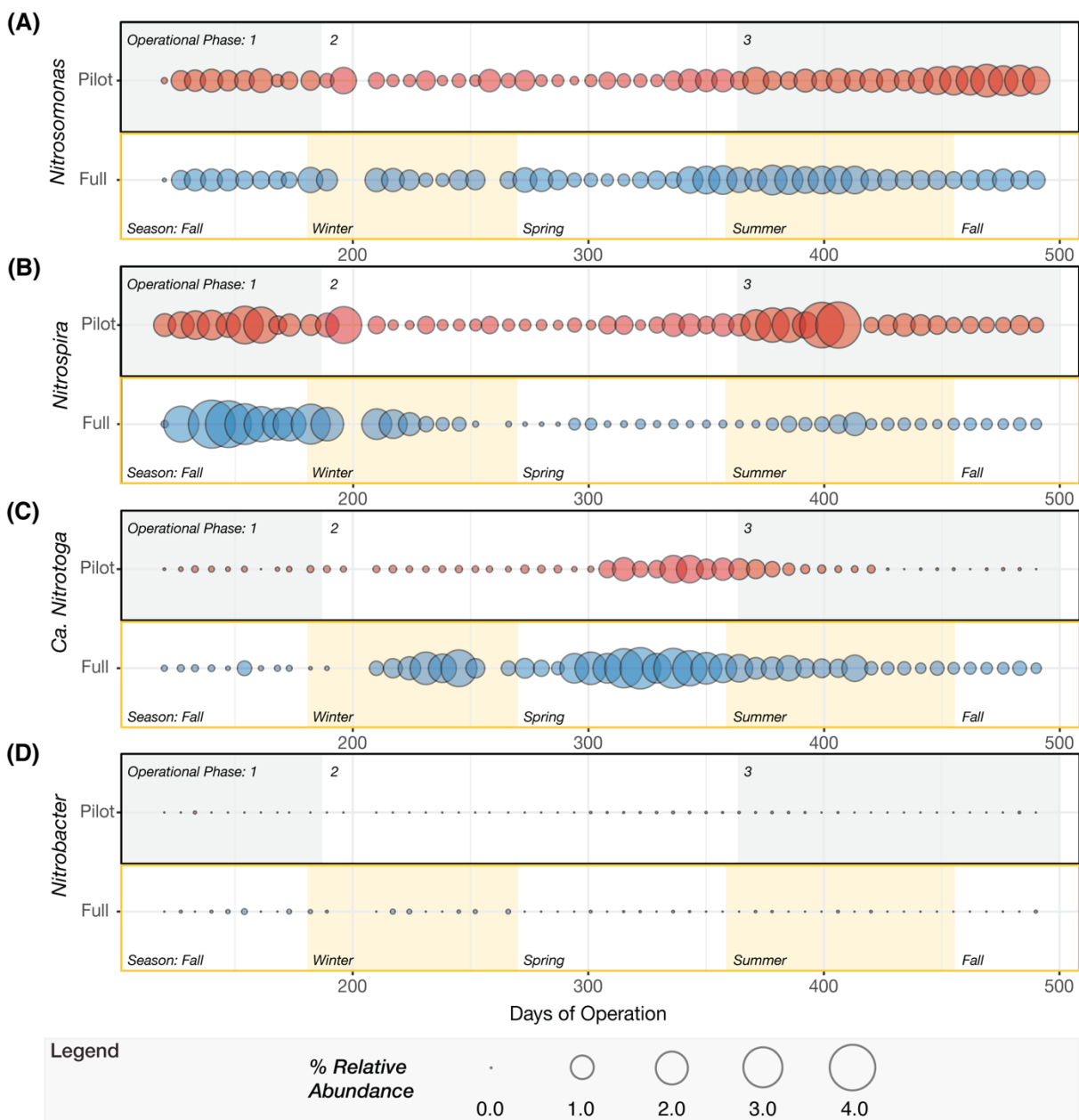


Figure 2-9. Balloon plots representing temporal variation in relative abundance of bacterial genera that participate in nitrification, (A) *Nitrosomonas*, (B) *Nitrospira*, (C) *Candidatus Nitrotoga*, and (D) *Nitrobacter*. For each genus, the pilot plant is described in the upper panel and full-scale plant in the lower panel. Relative abundance was determined based on the entire community. The shaded areas in the upper panels denote the three operational phases whereas the shaded areas in the lower panels indicate the meteorological seasons between days 120 and 490 of pilot plant operation.

Nitrosomonas sp. Nm86 (AY123798.1), denovo201 shared a 99.3% sequence identity with *Nitrosomonas* sp. JL21 (AB000700.1) and *Nitrosomonas* sp. Nm59 (AY123811.1), whereas denovo166 shared 99.6% sequence identity with *N. oligotropha* strain Nm45 (NR_104820.1).

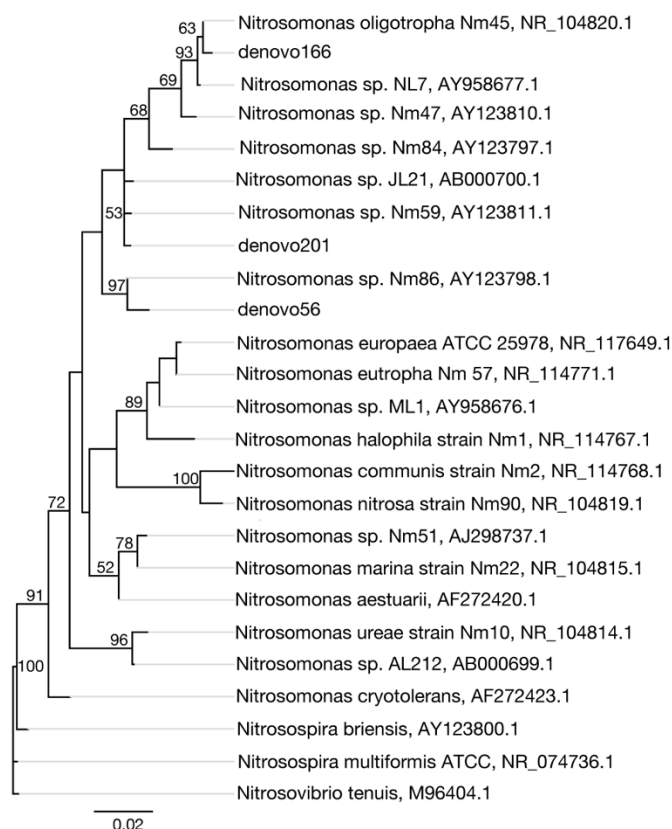


Figure 2-10. Neighbor-joining consensus tree generated from an alignment of published 16S rRNA sequences and sequences retrieved in this study, rooting with *Nitrosovibrio*. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. The scale bar indicates a 2% sequence difference. Accession numbers are presented after the sequence names.

Nitrospira was the most abundant nitrite oxidizing bacteria (NOB) in the pilot plant and was one of the top 30 core taxa but was not in the top 30 taxa in the full-scale plant (Figure 2-6). During the final operational phase, the mean relative abundance of *Nitrospira* was $1.30 \pm 1.2\%$ and $0.25 \pm 0.17\%$ in the pilot- and full-scale plants, respectively. In the full-scale plant, *Nitrospira* and *Candidatus Nitrotoga* (Ca.

Nitrotoga) had alternating abundances (Figure 2-9B and 2-9C). *Ca. Nitrotoga* was also detected in the pilot plant on a majority of the sampling days; however, this population was greater than 0.1% abundance only between days 308 and 392. In addition, *Nitrobacter* populations were not significant NOB to either plant (Figure 2-9D). Despite the increased abundance of *Nitrospira* in the pilot plant, the scattered distribution of all high-DO and low-DO sample points in the PCoA did not reveal a diverging NOB population in response to changes in DO (ANOSIM $P = 0.002$, $R = 0.18$, $n_{\text{otu}} = 29$) (Figure 2-8D). Again, although the p-value suggests that the differences are statistically significant, the R test statistic < 0.25 corresponds to “barely separable” groups [35]. One *Nitrospira* OTU (i.e. denovo17) was consistently the dominant *Nitrospira* in both plants (Figure S2-3). The 16S rRNA of this OTU had 100% sequence identity to *N. defluvii* (NR_074700.1) (Figure S2-4).

Energy Savings

Since an outcome of successfully operating a high-rate BNR process with minimal aeration is a reduction in energy, we performed an analysis of energy savings under the scenario of reducing DO at the Nine Springs WWTP. In this analysis, the current Nine Springs WWTP full-scale aeration energy usage was calculated and compared to energy usage in various DO scenarios.

The present-day full-scale operation uses aeration that results in averages of 0.9 mg O_2/L , 2.9 mg O_2/L and 4.3 mg O_2/L in the first, second, and third zones of the aerated tanks, respectively. The calculated aeration energy requirement to maintain these DO

concentrations at full-scale was 12,440,000 kWh/year (Table S2-3), which was within 1% of independent calculations described in a recent energy baseline study [38].

The DO scenarios used in this study (Figure 2-11) assumed one DO concentration across the entire aerated portion of the treatment train. If the DO were to be consistently controlled at 2 mg O₂/L, without any other changes or upgrades to the aeration system, a potential 10% aeration energy savings could be realized. Furthermore, if the plant were to operate with the lower DO conditions used in the pilot plant (e.g., maintaining an average DO concentration of 0.33 mg O₂/L, as was the case in the third phase of operation), nearly 25% in aeration energy savings could be achieved. This level of energy reduction correlates with savings of approximately \$262,000 in electricity costs per year (Table S2-3). This calculation assumes that the WWTP will maintain their current blower efficiency (63%) at the lower air flow rates. Additional savings could be realized from equipment upgrades, and if SND in the low-DO tanks could be stabilized and maintained at a predictable level.

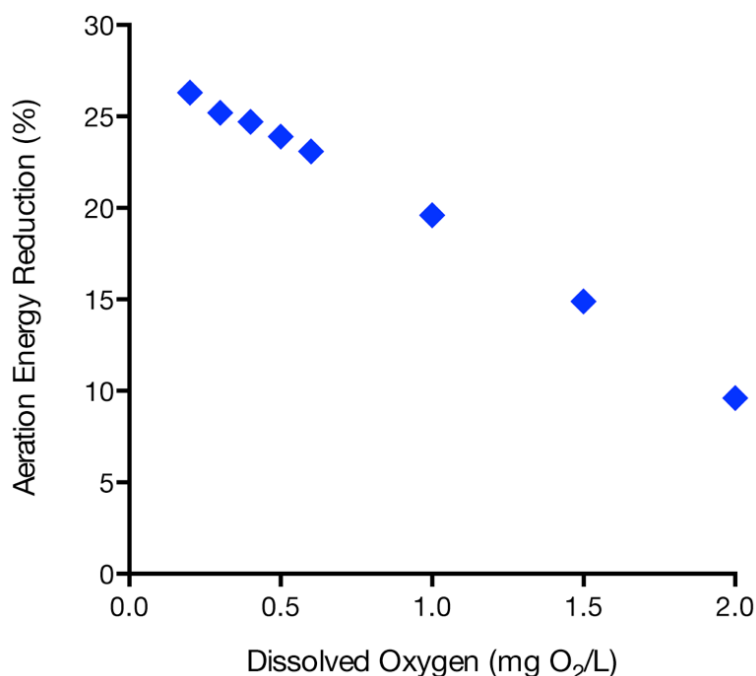


Figure 2-11. Estimated energy use reductions at the Nine Springs WWTP for 8 different DO scenarios.

Discussion

Achieving energy neutral or energy positive operation of a WWTP often requires a combination of increasing energy production and reducing energy consumption [39]. To increase energy production, facilities receive external organic waste to augment biogas generation [40, 41], use alternative sources of renewable energy, such as solar or wind power [42, 43], or redirect the majority of the wastewater organics to the anaerobic digester via enhanced primary settling [44-46]. Decreasing energy consumption relies on upgrading equipment to more energy efficient technology [1], reducing heat losses [43], and minimizing pumping of biosolids, liquid streams, and air supplies [47]. Some of these approaches can be constrained by the treatment objectives at the WWTP. For instance, redirecting organics to the anaerobic digester may not be possible if efficient

EBPR is required, as organics are needed to enrich for PAOs that efficiently perform the EBPR cycle.

A number of studies have shown that decreasing air delivery into the activated sludge basin is an effective approach to reduce energy use, especially in plants that do not have stringent P removal requirements, or plants that use chemical precipitation for P removal [7, 46]. However, there is a paucity of information regarding the performance of EBPR under reduced aeration, and more importantly, the effect of reduced aeration in BNR systems that require efficient and simultaneous biological nitrogen and P removal has only been studied at the bench-scale [13, 14]. Thus, to fill these knowledge gaps we used a pilot-scale system to gradually transform a high-rate and high-DO BNR process to high-rate and low-DO conditions. The pilot-scale plant received primary effluent from the adjoining full-scale plant (Nine Springs WWTP, Madison, WI) and experienced the same temporal variability in chemical composition and temperature as the full-scale plant.

During 16 months of operation, aeration was gradually reduced in the pilot plant until the average DO was 0.33 mg O₂/L. Over this entire operational period, which included one winter season (Figure 2-2), there was stable EBPR performance and nearly 90% P removal. In addition, once low-DO conditions were reached, nitrification efficiency was maintained with nearly 70% of the nitrogen denitrified without the need for internal recycling of sludge. Throughout low-DO operation, the average TKN removal efficiency was greater than 96%, not including days of mechanical disruption that temporarily decreased nitrification performance. EBPR performance was not impacted by these

mechanical disruptions, and complete nitrification was regained by temporarily discontinuing normal wasting procedures.

Given the efficient performance of the pilot plant, we carried out additional experiments to evaluate the microbial communities enriched in the plant and to elucidate factors that make low-DO high-rate BNR possible. Regarding EBPR efficiency, the collective results suggest that there were no major changes on the PAO performing EBPR as the plant underwent the transition from high to low-DO. Interestingly, the kinetic experiments showed that the PAO present at high-DO already had high affinity for oxygen (Figure 2-4), and after low-DO operation in the pilot plant, the observed oxygen half saturation constant for PAO ($K_{DO} = 0.09 \text{ mg O}_2/\text{L}$) was remarkably low. This observed half saturation constant is nearly 70% less than the value determined for a *Ca. Accumulibacter* enrichment ($K_{DO} = 0.27 \text{ mg O}_2/\text{L}$) [15], and 55% less than the typical value used in activated sludge modeling ($K_{DO} = 0.20 \text{ mg O}_2/\text{L}$, ASM2d) [48]. Even after long-term low-DO operation, results from the PCoA analysis (Figure 2-8) did not reveal any divergence in *Ca. Accumulibacter* in the two plants and the 16S rRNA tag sequencing revealed the predominance of two *Ca. Accumulibacter* OTUs that remained dominant through the three operational phases (Figure S2-1). The only observation that may reflect an adaptation to low-DO is the higher specific P uptake rate observed in the pilot plant (Figure 2-4).

The possibility of denitrifying PAO (DPAO) being enriched in low-DO BNR operation has been proposed [14], although the identity of *Ca. Accumulibacter* able to denitrify remains controversial [14, 49-52]. There also remains uncertainty about the preferred electron acceptor, with some reports suggesting that certain DPAO can only reduce

nitrite [52, 53], others suggesting that nitrate is preferentially used [54, 55], and more recently, that some DPAO may simultaneously use oxygen and oxidized forms of nitrogen as electron acceptors in low-DO conditions [14]. Since Camejo et al. demonstrated the greatest P uptake when nitrate and oxygen were simultaneously present, it is possible these PAO may have contributed to SND within the aerobic zone of the pilot plant [14]. However, the extent of SND in the pilot plant was measured, but not found to be stable (Figure 2-2). Furthermore, the measured capacity of PAO to denitrify (Figure 2-3) in the pilot plant did not support a high enrichment of DPAO. Overall, these results are consistent with PAOs with high affinity for oxygen being responsible for most of the P removal at low-DO conditions.

In regard to low-DO nitrification, the experimental results suggest a change in the AOB community in response to decreasing DO. The kinetic experiments showed an increase in observed oxygen affinity (Figure 2-4). The measured increase in oxygen affinity may be influenced by a reduction in mass transfer limitations due to smaller flocs and smaller AOB microcolonies (Figure 2-5), as well as the change in the predominant AOB. The PCoA analysis describes a diverging *Nitrosomonas* community (Figure 2-8), and the OTU-level analysis reveals a transition from a dominant *Nitrosomonas* OTU at high-DO conditions to two other OTUs at low-DO (Figure S2-2). The identification of *Nitrosomonas* as the most abundant AOB in the full- and pilot-scale plants is consistent with this AOB being prevalent in wastewater treatment even under low-DO conditions [56-58]. In particular, the enrichment of an *N. oligotropha*-related OTU (Figure 2-10) agrees with this AOB lineage having high affinity for oxygen [56, 58, 59]. The absence of AOA, which have been described as important in low-DO oxidation ditch reactors

[60], but not found in low-DO enrichments from the Nine Springs WWTP [10] supports the hypothesis that AOA are not needed to achieve efficient low-DO nitrification in BNR systems [10].

With the level of resolution afforded by the 16S rRNA tag sequencing, we found no evidence of NOB variations as a result of low-DO (Figure 2-8), since the *Nitrospira* abundance in the reactor remained relatively constant throughout the pilot plant operation. The 16S rRNA partial sequence of the abundant NOB OTU had 100% identity to *N. defluvii*. This result is consistent with earlier evidence that NOB related to *N. defluvii* are able to successfully perform nitrite oxidation at low-DO conditions [61].

Overall, these results give confidence that high-rate BNR systems can be successfully operated with low-DO, confirming that this approach is suitable for reducing energy consumption at WWTPs. We estimated that reducing DO at the Nine Springs WWTP to the conditions used in the pilot plant could represent an approximate 25% reduction in energy use without compromising effluent quality. This reduction would mostly be achieved by the increase in oxygen transfer efficiency from the larger DO gradient generated when maintaining low-DO, since we did not take into consideration the reduction in oxygen requirements due to improved denitrification nor any potential change in the standard oxygen transfer efficiency (see Supplementary Document). Further reductions in energy use could theoretically be accomplished by aeration equipment upgrades. Although these calculations are specific to the Nine Springs WWTP, they reflect the likely scenario of many WWTPs with aging infrastructure that first implemented high-rate BNR decades ago. For such plants, retrofitting to high-rate

low-DO BNR may be an attractive option that contributes to the quest for energy neutrality during wastewater treatment.

Conclusions

- Slow step-wise oxygen reductions were identified as a strategy to transition a conventional high-rate high-DO BNR process to high-rate low-DO BNR operation.
- Successful low-DO nitrification combined with low-DO P removal was demonstrated at the pilot-scale, with average DO concentrations as low as 0.33 mg O₂/L and average effluent TKN, TAN, and P concentrations at 1.4 ± 0.30 mg TKN/L, 0.13 ± 0.11 mg TAN/L, and 0.77 ± 0.62 mg PO₄³⁻-P/L, respectively. Furthermore, TKN, TAN, and P removal efficiencies at low-DO were an average of 96%, 99%, and 89%, respectively.
- Adaptation of nitrifiers to low-DO conditions resulted in an observed higher affinity for oxygen ($K_{DO}=0.30$ mg O₂/L) compared to high-DO nitrification ($K_{DO}=1.38$ mg O₂/L), which may have been attributed to a decrease in mass transfer limitations in the smaller floc and AOB microcolonies (Figure 2-5), and a change in the dominant AOB population (Figure S2-2).
- PAO in the high-DO full-scale plant exhibited a remarkably high affinity for oxygen. Adaptation of PAO to low-DO conditions was indicated through much higher specific rate of P uptake when compared to PAO from the full-scale plant.
- Up to an estimated 25% decrease in aeration energy requirements may be achieved if the full-scale plant converts from high-rate high-DO BNR to high-rate low-DO BNR. Additional reductions could be achieved with upgrades to high efficiency aeration equipment.

Acknowledgements

This work was partially supported by funding from the National Science Foundation (CBET-1435661) and the Madison Metropolitan Sewerage District. This work is also supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256259. We thank Jackie Bastyr-Cooper for her assistance in the lab with training, protocols and equipment, and Rodolfo Perez for his help setting up the DO monitoring and control system. We greatly appreciate Alexandria Sabarots for her support with the batch kinetic experiments and Elizabeth Erb for her dedication to collect and record detailed microcolony size data. We also thank Bradley Glisczinski, Tyler Gates, Brian Owen, Sam Acker, Diana Barrera, Matthew Dysthe, Rahim Ansari, Cassie Ganos, Kate Dapolito, and Randy Abilmona for support with laboratory analysis and reactor operation, and Michael Elk for technical support with next-generation 16S rRNA tag sequencing. The U.S. Environmental Protection Agency, through its Office of Research and Development, partially funded and collaborated in the research described herein. Any opinions expressed in this paper are those of the authors and do not reflect the views of the agency; therefore, no official endorsement should be inferred. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

1. Appelbaum, B., *Water and Sustainability: U.S. Electricity Consumption for Water Supply & Treatment - The Next Half Century*, in *Water and Sustainability*. 2002, EPRI: Palo Alto, CA. <http://www.circleofblue.org/wp-content/uploads/2010/08/EPRI-Volume-4.pdf>
2. Daigger, G.T. and H.X. Littleton, *Simultaneous biological nutrient removal: a state-of-the-art review*. *Water Environ Res*, 2014. **86**(3): p. 245-57. 10.2175/106143013x13736496908555
3. Bertanza, G., *Simultaneous nitrification-denitrification process in extended aeration plants: Pilot and real scale experiences*. *Water Science And Technology*, 1997. **35**(6): p. 53-61. Doi 10.1016/S0273-1223(97)00095-4
4. Park, H.D., J.M. Regan, and D.R. Noguera, *Molecular analysis of ammonia-oxidizing bacterial populations in aerated-anoxic orbital processes*. *Water Sci Technol*, 2002. **46**(1-2): p. 273-80. <https://www.ncbi.nlm.nih.gov/pubmed/12216636>
5. Giraldo, E., et al. *Ammonia Oxidizing Archaea, AOA, Population and Kinetic Changes in a Full Scale Simultaneous Nitrogen and Phosphorus Removal MBR*. in *Proceedings of the Water Environment Federation, WEFTEC*. 2011. Los Angeles, CA: Water Environment Federation. 10.2175/193864711802721596
6. Littleton, H., et al. *Characterizing Activity of Deammonification, Nitrification, denitrification by AOB, AOA, NOB, Anammox and Heterotrophic denitrifier in Full Scale NPXpress MBR Plants*. in *Proceedings of the Water Environment Federation, WEFTEC*. 2013. Chicago, IL: Water Environment Federation.
7. Amand, L. and B. Carlsson, *Optimal aeration control in a nitrifying activated sludge process*. *Water Res*, 2012. **46**(7): p. 2101-10. 10.1016/j.watres.2012.01.023

8. Amand, L., G. Olsson, and B. Carlsson, *Aeration control - a review*. Water Sci Technol, 2013. **67**(11): p. 2374-98. 10.2166/wst.2013.139
9. Rieger, L., et al., *Ammonia-based feedforward and feedback aeration control in activated sludge processes*. Water Environ Res, 2014. **86**(1): p. 63-73. 10.2175/106143013x13596524516987
10. Fitzgerald, C.M., et al., *Ammonia-oxidizing microbial communities in reactors with efficient nitrification at low-dissolved oxygen*. Water Res, 2015. **70**: p. 38-51. 10.1016/j.watres.2014.11.041
11. Liu, G. and J. Wang, *Long-term low DO enriches and shifts nitrifier community in activated sludge*. Environ Sci Technol, 2013. **47**(10): p. 5109-17. 10.1021/es304647y
12. Li, H., Y. Chen, and G. Gu, *The effect of propionic to acetic acid ratio on anaerobic-aerobic (low dissolved oxygen) biological phosphorus and nitrogen removal*. Bioresource Technology, 2008. **99**(10): p. 4400-4407. <https://doi.org/10.1016/j.biortech.2007.08.032>
13. Zheng, X., et al., *The investigation of effect of organic carbon sources addition in anaerobic-aerobic (low dissolved oxygen) sequencing batch reactor for nutrients removal from wastewaters*. Bioresource Technology, 2009. **100**(9): p. 2515-2520. <https://doi.org/10.1016/j.biortech.2008.12.003>
14. Camejo, P.Y., et al., *Candidatus Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors*. Water Res, 2016. **102**: p. 125-137. 10.1016/j.watres.2016.06.033
15. Carvalheira, M., et al., *The impact of aeration on the competition between polyphosphate accumulating organisms and glycogen accumulating organisms*. Water Res, 2014. **66**: p. 296-307. 10.1016/j.watres.2014.08.033

16. Rice, E.W., et al., eds. *Standard Methods for the Examination of Water and Wastewater*. 22 ed. 2012, American Public Health Association, American Water Works Association, Water Environment Federation: Washington, D.C., <https://www.e-wef.org/Default.aspx?TabID=251&productId=17997>
17. Monod, J., *The growth of bacterial cultures*. Annual Review of Microbiology, 1949. 3(1): p. 371-394. 10.1146/annurev.mi.03.100149.002103
18. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms*. ISME J, 2012. 6(8): p. 1621-4. 10.1038/ismej.2012.8
19. Joshi, N. and J. Fass, *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]*. 2011. <https://github.com/najoshi/sickle>.
20. Magoc, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. 27(21): p. 2957-63. 10.1093/bioinformatics/btr507
21. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data*. Nat Methods, 2010. 7(5): p. 335-6. 10.1038/nmeth.f.303
22. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST*. Bioinformatics, 2010. 26(19): p. 2460-1. 10.1093/bioinformatics/btq461
23. McIlroy, S.J., et al., *MiDAS: the field guide to the microbes of activated sludge*. Database (Oxford), 2015. 2015: p. bav062. 10.1093/database/bav062
24. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data*. PLoS One, 2013. 8(4): p. e61217. 10.1371/journal.pone.0061217

25. R Core Team, *R: A language and environment for statistical computing*. 2016, R Foundation for Statistical Computing: Vienna, Austria. <https://www.R-project.org/>
26. Bray, J.R. and J.T. Curtis, *An Ordination of the Upland Forest Communities of Southern Wisconsin*. Ecological Monographs, 1957. 27(4): p. 325-349. 10.2307/1942268
27. Clarke, K.R., *Non-parametric multivariate analyses of changes in community structure*. Australian Journal of Ecology, 1993. 18(1): p. 117-143. 10.1111/j.1442-9993.1993.tb00438.x
28. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2009, New York: Springer-Verlag. 10.1007/978-0-387-98141-3
29. Warnes, G.R., et al., *gplots: Various R Programming Tools for Plotting Data*. 2016. <https://CRAN.R-project.org/package=gplots>
30. Johnson, M., et al., *NCBI BLAST: a better web interface*. Nucleic Acids Res, 2008. 36(Web Server issue): p. W5-9. 10.1093/nar/gkn201
31. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data*. Bioinformatics, 2012. 28(12): p. 1647-9. 10.1093/bioinformatics/bts199
32. Picioreanu, C., J. Perez, and M.C. van Loosdrecht, *Impact of cell cluster size on apparent half-saturation coefficients for oxygen in nitrifying sludge and biofilms*. Water Res, 2016. 106: p. 371-382. 10.1016/j.watres.2016.10.017
33. Mann, H.B. and D.R. Whitney, *On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other*. The Annals of Mathematical Statistics, 1947. 18(1): p. 50-60. <http://www.jstor.org/stable/2236101>

34. Nguyen, H.T., et al., *High diversity and abundance of putative polyphosphate-accumulating Tetrasphaera-related bacteria in activated sludge systems*. FEMS Microbiol Ecol, 2011. **76**(2): p. 256-67. 10.1111/j.1574-6941.2011.01049.x
35. Ramette, A., *Multivariate analyses in microbial ecology*. FEMS Microbiol Ecol, 2007. **62**(2): p. 142-60. 10.1111/j.1574-6941.2007.00375.x
36. Garcia Martin, H., et al., *Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities*. Nat Biotechnol, 2006. **24**(10): p. 1263-9. 10.1038/nbt1247
37. McMahon, K.D., et al., *Polyphosphate kinase genes from full-scale activated sludge plants*. Appl Microbiol Biotechnol, 2007. **77**(1): p. 167-73. 10.1007/s00253-007-1122-6
38. Schroedel Jr., R. and R.A. Wirtz, *Energy Baseline and Optimization Roadmap Study*, R. Schroedel Jr. and R.A. Wirtz, Editors. 2014, Brown and Caldwell in association with Strand Associates. p. 1-195. <http://www.madsewer.org/Portals/0/Planning/Sustainability/Final%20Report%20Energy%20Study%20-%20Complete.pdf>
39. Kroiss, H. and Y. Cao, *Energy considerations*, in *Activated Sludge - 100 Years and Counting*, D. Jenkins and J. Wanner, Editors. 2014, IWA Publishing: London. p. 221-244. 10.2166/9781780404943
40. Gómez, X., et al., *Anaerobic co-digestion of primary sludge and the fruit and vegetable fraction of the municipal solid wastes*. Renewable Energy, 2006. **31**(12): p. 2017-2024. 10.1016/j.renene.2005.09.029
41. Gao, H., Y.D. Scherson, and G.F. Wells, *Towards energy neutral wastewater treatment: methodology and state of the art*. Environ Sci Process Impacts, 2014. **16**(6): p. 1223-46. 10.1039/c4em00069b

42. Argaw, N., *Renewable Energy in Water and Wastewater Treatment Applications*. 2003, National Renewable Energy Laboratory. <http://www.nrel.gov/docs/fy03osti/30383.pdf>
43. Chae, K.-J. and J. Kang, *Estimating the energy independence of a municipal wastewater treatment plant incorporating green energy resources*. *Energy Conversion and Management*, 2013. **75**: p. 664-672. 10.1016/j.enconman.2013.08.028
44. Jenicek, P., et al., *Energy self-sufficient sewage wastewater treatment plants: is optimized anaerobic sludge digestion the key?* *Water Sci Technol*, 2013. **68**(8): p. 1739-44. 10.2166/wst.2013.423
45. Remy, C., M. Boulestreau, and B. Lesjean, *Proof of concept for a new energy-positive wastewater treatment scheme*. *Water Sci Technol*, 2014. **70**(10): p. 1709-16. 10.2166/wst.2014.436
46. Zaborowska, E., K. Czerwionka, and J. Makinia, *Strategies for achieving energy neutrality in biological nutrient removal systems - a case study of the Slupsk WWTP (northern Poland)*. *Water Science and Technology*, 2016. 10.2166/wst.2016.564
47. Hoppock, D. and M. Webber. *Energy needs and opportunities at POTWs in the United States*. in *Proceedings of the American Society of Mechanical Engineers (ASME) 2nd International Conference on Energy Sustainability*. 2008. Jacksonville, FL.
48. Henze, M., et al., *Activated Sludge Models ASM1, ASM2, ASM2d and ASM3*. 2000, IWA Publishing: London, UK. doi:10.1007/s13398-014- 0173-7.2
49. Zeng, R.J., et al., *Simultaneous nitrification, denitrification, and phosphorus removal in a lab-scale sequencing batch reactor*. *Biotechnol Bioeng*, 2003. **84**(2): p. 170-8. 10.1002/bit.10744

50. Flowers, J.J., et al., *Denitrification capabilities of two biological phosphorus removal sludges dominated by different "Candidatus Accumulibacter" clades*. Environ Microbiol Rep, 2009. 1(6): p. 583-588. 10.1111/j.1758-2229.2009.00090.x
51. Oehmen, A., et al., *Assessing the abundance and activity of denitrifying polyphosphate accumulating organisms through molecular and chemical techniques*. Water Sci Technol, 2010. 61(8): p. 2061-8. 10.2166/wst.2010.976
52. Kim, J.M., et al., *Characterization of the Denitrification-Associated Phosphorus Uptake Properties of "Candidatus Accumulibacter phosphatis" Clades in Sludge Subjected to Enhanced Biological Phosphorus Removal*. Applied and Environmental Microbiology, 2013. 79(6): p. 1969-1979. 10.1128/aem.03464-12
53. Guisasola, A., et al., *Failure of an enriched nitrite-DPAO population to use nitrate as an electron acceptor*. Process Biochemistry, 2009. 44(7): p. 689-695. 10.1016/j.procbio.2009.02.017
54. Hu, J.Y., et al., *A new method for characterizing denitrifying phosphorus removal bacteria by using three different types of electron acceptors*. Water research, 2003. 37(14): p. 3463-71. 10.1016/S0043-1354(03)00205-7
55. Jiang, Y., et al., *Dynamic Response of Denitrifying Poly-P Accumulating Organisms Batch Culture to Increased Nitrite Concentration as Electron Acceptor*. Journal of Environmental Science and Health, Part A, 2006. 41(11): p. 2557-2570. 10.1080/10934520600927930
56. Bellucci, M., et al., *Low-dissolved-oxygen nitrifying systems exploit ammonia-oxidizing bacteria with unusually high yields*. Appl Environ Microbiol, 2011. 77(21): p. 7787-96. 10.1128/AEM.00330-11
57. Park, H.D. and D.R. Noguera, *Evaluating the effect of dissolved oxygen on ammonia-oxidizing bacterial communities in activated sludge*. Water Res, 2004. 38(14-15): p. 3275-86. 10.1016/j.watres.2004.04.047

58. Park, H.D. and D.R. Noguera, *Characterization of two ammonia-oxidizing bacteria isolated from reactors operated with low dissolved oxygen concentrations*. J Appl Microbiol, 2007. **102**(5): p. 1401-17. 10.1111/j.1365-2672.2006.03176.x
59. Gieseke, A., et al., *Community structure and activity dynamics of nitrifying bacteria in a phosphate-removing biofilm*. Appl Environ Microbiol, 2001. **67**(3): p. 1351-62. 10.1128/AEM.67.3.1351-1362.2001
60. Park, H.D., et al., *Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors*. Appl Environ Microbiol, 2006. **72**(8): p. 5643-7. 10.1128/AEM.00402-06
61. Park, H.D. and D.R. Noguera, *Nitrospira community composition in nitrifying reactors operated with two different dissolved oxygen levels*. J Microbiol Biotechnol, 2008. **18**(8): p. 1470-4.
<https://www.ncbi.nlm.nih.gov/pubmed/18756110>

Supplementary Tables

Table S2-1. Average operational parameters for the pilot- and full-scale plants

Plant	Average DO Concentrations (mg O ₂ /L)			Solids Retention Time (days)	Total Hydraulic Retention Time (hours)	Aerobic Hydraulic Retention Time (hours)
	Tank 1	Tank 2	Tank 3			
Pilot phase 1	0.51 ± 0.28	0.72 ± 0.44	1.1 ± 0.28	10 ± 3.4	9.0 - 10	5.7 - 6.4
Pilot phase 2	0.41 ± 0.20	0.54 ± 0.30	0.67 ± 0.30			
Pilot phase 3	0.23 ± 0.13	0.48 ± 0.12	0.28 ± 0.25			
Full-scale, Plant 3	Basin 1	Basin 2	Basin 3	10	17	13
	0.40	0.83	4.20			

Table S2-2. 16S rRNA-targeted oligonucleotide probes used in this study

Oligonucleotide Probe	Sequence of probe (5'-3')	Targeted Organism	Reference
NS01225	CGCCATTGTATTACGTGTGA	All ammonia-oxidizing bacteria within the β -subgroup of <i>Proteobacteria</i>	[1]
EUB338	GCTGCCTCCCGTAGGAGT	Domain Bacteria	[2]
EUB338-II	GCAGCCACCCGTAGGTGT	<i>Planctomycetales</i>	[3]
EUB338-III	GCTGCCACCCGTAGGTGT	<i>Verrucomicrobiales</i> <i>Bacteria</i>	[3]

Table S2-3. Energy savings summary table, including modeled basin DO, calculated total energy usage, percent energy savings and change in annual energy costs based on \$0.0835/kWh. Shaded rows represent the current scenario and the low-DO scenario that corresponds to the stable low-DO operation of the pilot plant.

DO Concentrations (mg O ₂ /L)			Calculated Total Energy Use (kWh/yr)	Percent Aeration Energy Reduction (%)	Change in Energy Use (kWh/yr)	Change in Annual Energy Costs (\$)
Basin 1	Basin 2	Basin 3				
0.90	2.90	4.30	12,400,000	0.0%	0	\$0
2.00	2.00	2.00	11,200,000	9.6%	-1,190,000	-\$99,700
1.50	1.50	1.50	10,600,000	14.9%	-1,850,000	-\$155,000
1.00	1.00	1.00	10,000,000	19.6%	-2,440,000	-\$204,000
0.60	0.60	0.60	9,570,000	23.1%	-2,870,000	-\$240,000
0.50	0.50	0.50	9,470,000	23.9%	-2,970,000	-\$248,000
0.40	0.40	0.40	9,370,000	24.7%	-3,070,000	-\$256,000
0.33	0.33	0.33	9,300,000	25.2%	-3,140,000	-\$262,000
0.20	0.20	0.20	9,170,000	26.3%	-3,270,000	-\$273,000

Supplementary Figures

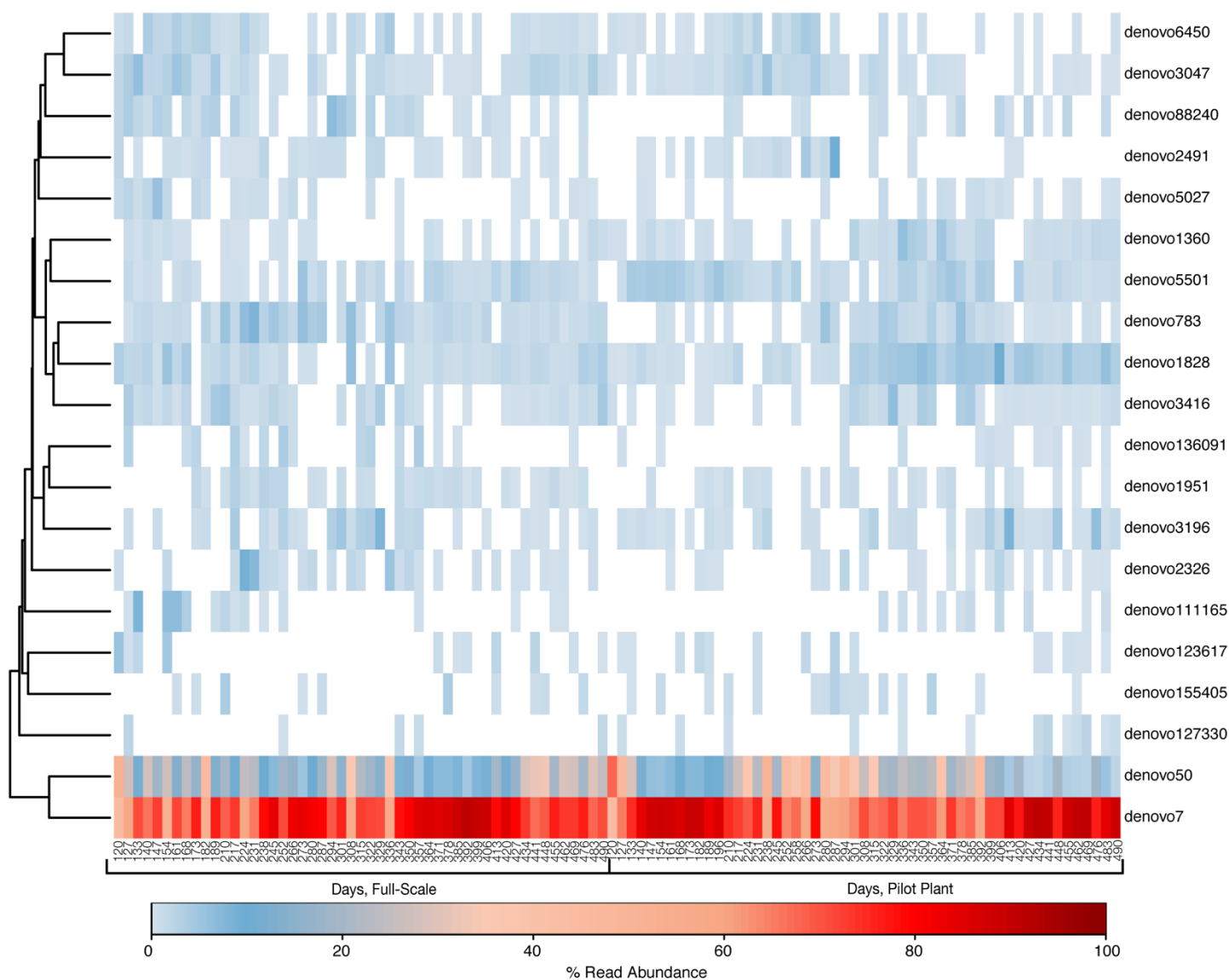


Figure S2-1. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Candidatus Accumulibacter* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Ca. Accumulibacter* OTUs, not the whole community.

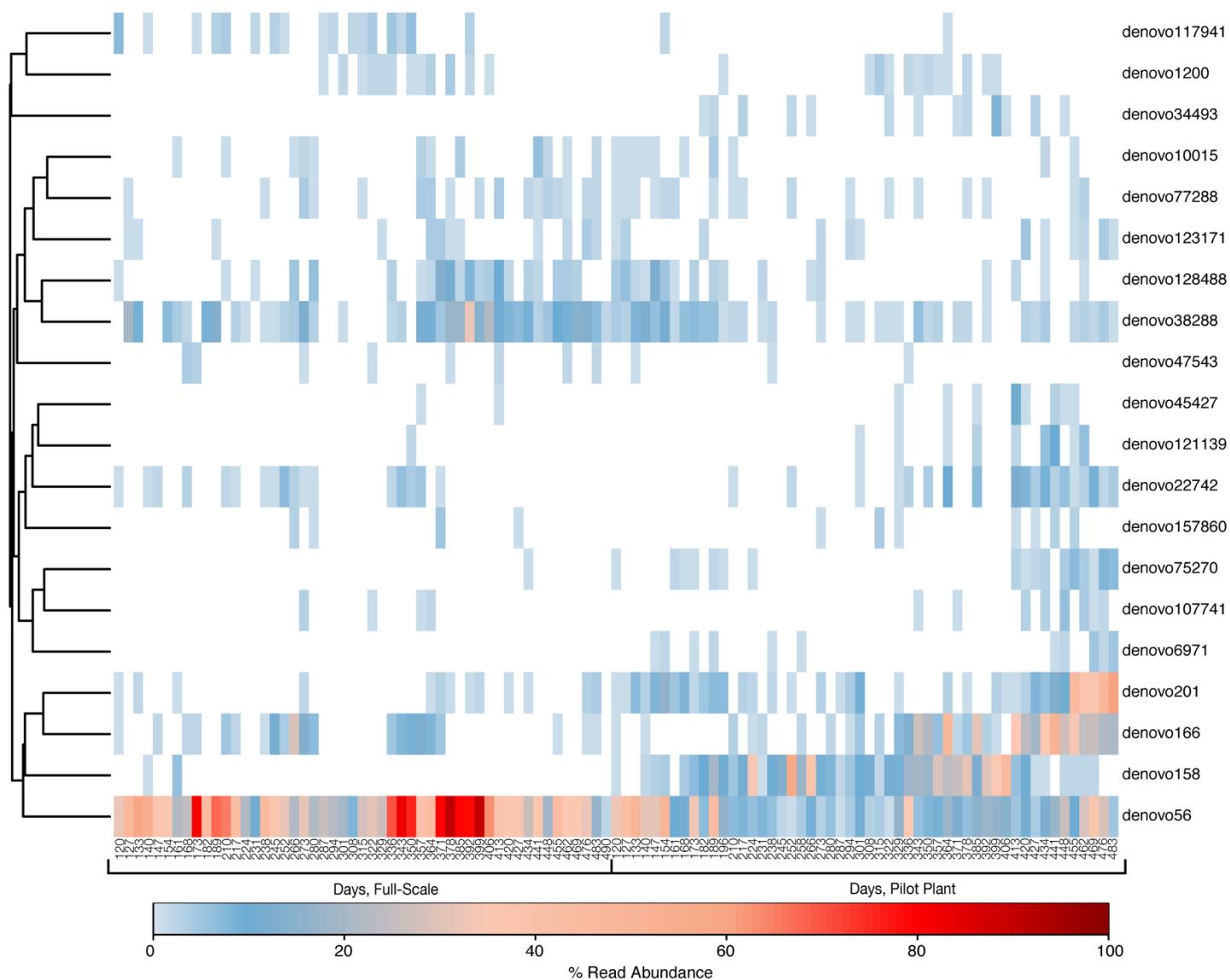


Figure S2-2. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Nitrosomonas* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Nitrosomonas* OTUs, not the whole community.

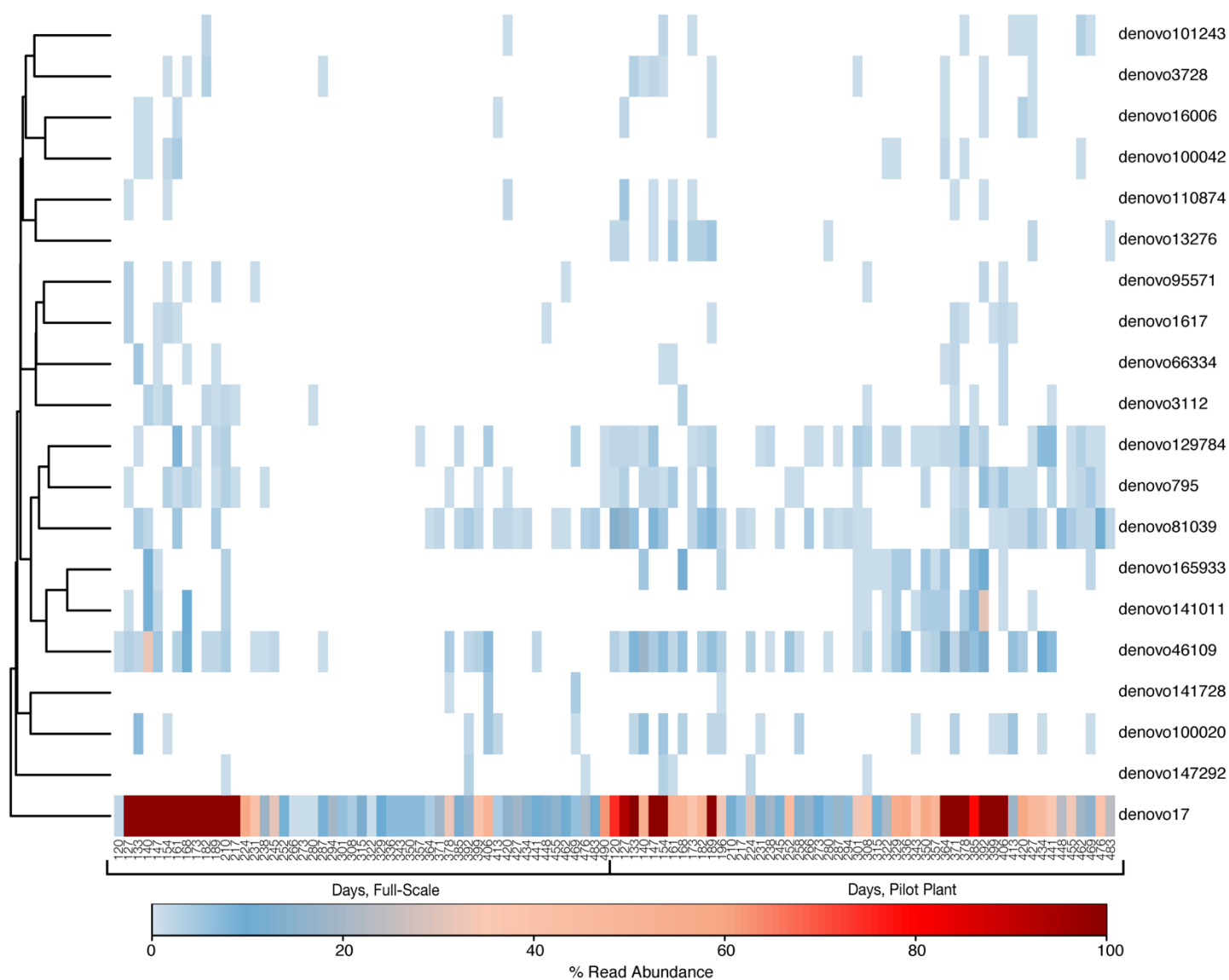


Figure S2-3. Heatmap dendrogram demonstrating changes in the relative abundance of OTUs classified as *Nitrospira* in the full-scale and pilot-scale plants. Note that the read abundance was calculated using the total number of reads in the subset of *Nitrospira* OTUs, not the whole community.

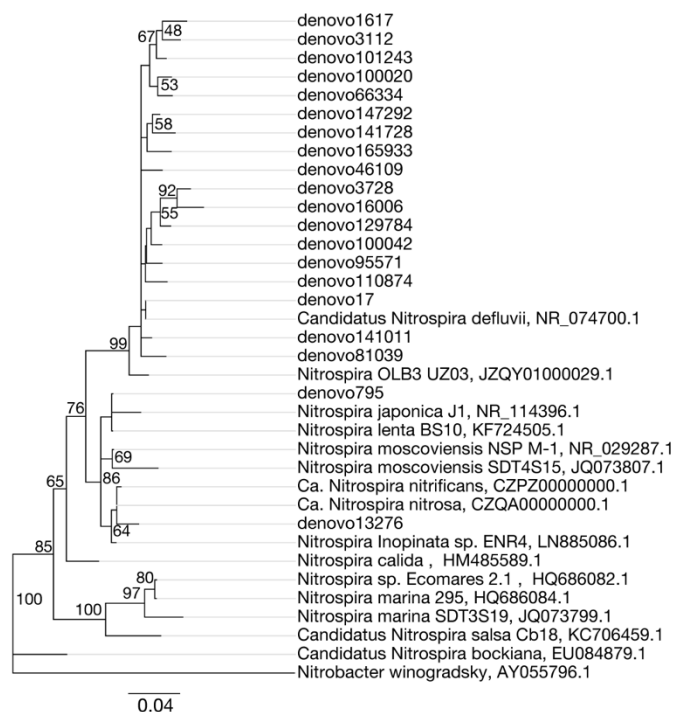


Figure S2-4. Neighbor-joining consensus tree generated from an alignment of published 16S rRNA sequences and sequences retrieved in this study. *Nitrobacter winogradskyi* was used to root the tree. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. The scale bar indicates a 4% sequence difference. Accession numbers are presented after the sequence names.

Supplementary Methods

Floc and microcolony sizes

To compare floc from the full-scale plant operated with high DO and the pilot plant after stable low DO operation, grab samples were collected from each system, the biomass embedded in an agarose solution, and visualized via microscopy. Initially, an agarose solution was prepared by dissolving 1.5 grams of Multi-Purpose Agarose (Thermo Fisher Scientific, Waltham, MA) in 100mL of deionized water, and heated in a microwave on high until clear, approximately one minute. The heated solution was immediately filtered through a 0.45- μ m membrane filter (Nitrocellulose Membrane Filters, EMD Millipore Corp., Darmstadt, Germany) to remove impurities, and the filtrate stored at 50° C during sample preparation. The biomass samples were stained with NucBlue® Live ReadyProbes® Reagent following the manufacturer's protocol (Thermo Fisher Scientific, Waltham, MA). Two to three drops of activated sludge were carefully transferred to a petri dish using a wide mouth pipet and approximately 5 mL of the agarose solution was added. The petri dish was gently rotated to ensure an even distribution of floc and the agarose allowed to solidify. The petri dishes were immediately visualized with a Zeiss Microscope Axio Imager.Z2 and AxioCam MRM camera (Zeiss, Oberkochen, Germany) using the DAPI filter, and images were captured with the AxioVision Rel. 4.8 Software for Image Acquisition and Management for Light Microscopy (Zeiss, Oberkochen, Germany). The surface area of individual floc was measured using the threshold, trace, and analyze features within the ImageJ software, which accounted for the entire irregular shape [4]. An equivalent diameter was

calculated from the surface area of 356 floc and 373 floc originating from the full-scale and pilot plant, respectively.

Fluorescence *in situ* hybridization (FISH) was used to analyze AOB microcolonies from the full-scale and pilot plants. Biomass samples were collected from both the full-scale and pilot plant on day 371 and day 392 (during the third operational phase). The biomass was fixed using a 4% paraformaldehyde solution and stored at -20°C in a 1:1 volume of phosphate-buffered saline and absolute ethanol [5]. Fixed cells were placed on gelatin-coated glass slides and hybridized with a mixture of hybridization buffer and FISH probes. Oligonucleotide probes NSO1225 and an EUB mix (EUB338, EUB338-II, and EUB338-III) were used to detect AOB and Bacteria, respectively (Table S2-2). Microbial cells were counterstained with a 1 µg/ml of 4'6-diamidino-2-phenylindol (DAPI). Controls were performed to examine autofluorescence. The slides were visualized with a Zeiss Microscope Axio Imager.Z2 and AxioCam MRM camera (Zeiss, Oberkochen, Germany). Images were captured and analyzed with the AxioVision Rel. 4.8 Software for Image Acquisition and Management for Light Microscopy (Zeiss, Oberkochen, Germany). A positive signal for the NSO1225 oligonucleotide probe (Table S2-2) was used to identify 96 and 145 AOB microcolonies originating from the full-scale and pilot plant, respectively. An equivalent diameter for each microcolony was calculated from the maximum dimensions in the horizontal and vertical planes.

Calculations of oxygen requirement and energy savings

To estimate the energy savings associated with low-DO operation for the full-scale Nine Springs WWTP, oxygen transfer and oxygen requirement calculations were performed.

All calculations assumed one treatment train with a three-pass aerobic zone. The first, second, and third aeration basins are 1/7, 3/7, and 3/7 of the total aerated volume, respectively.

Initially, the oxygen requirement for the biodegradation of carbonaceous material plus the oxygen requirement for oxidation of ammoniacal nitrogen were calculated according to Equations S.1 and S.2 [6].

$$R_O = Q(S_O - S) - 1.42P_{X,bio} + 4.33Q(TKN) \quad (S2.1)$$

Where:

R_O = total oxygen required, grams per day (g/d)

Q = influent flow, meters³/day (m³/d), assumed 40 MGD (151,416 m³/d)

S_O = influent substrate concentration (BOD), (mg/L), assumed 150 mg BOD/L

S = effluent substrate concentration (BOD), (mg/L), assumed 0 mg BOD/L

$P_{X,bio}$ = biomass as VSS wasted, from Equation S.2 (g/d)

TKN = influent nitrogen that is oxidized, (mg TKN/L), assumed 35 mg TKN/L

$$P_{X,bio} = \frac{QY(S_O - S)\left(\frac{1kg}{10^3g}\right)}{1 + (k_d)SRT} + \frac{(f_d)(k_d)QY(S_O - S)SRT\left(\frac{1kg}{10^3g}\right)}{1 + (k_d)SRT} + \frac{QY_n(TKN)\left(\frac{1kg}{10^3g}\right)}{1 + (k_{dn})SRT} \quad (S.2)$$

Where:

Y = biomass yield, g of VSS per g of BOD consumed, assumed 0.5 g VSS/g BOD

Y_n = biomass yield for nitrogen, g VSS per g of $\text{NH}_4\text{-N}$, assumed 0.12 g VSS/g $\text{NH}_4\text{-N}$

k_d = endogenous decay coefficient for heterotrophic organisms, days^{-1} , assumed 0.12 d^{-1}

k_{dn} = endogenous decay coefficient for nitrifying organisms, days^{-1} , assumed 0.08 d^{-1}

f_d = fraction of cell mass remaining as cell debris, unitless, assumed 0.15

SRT = solids retention time, days (d), assumed 10 d

Subsequently, the total biological oxygen requirement was converted to pounds of oxygen per day ($\text{lb O}_2/\text{d}$) by multiplying R_o by 0.0022 lb/g . The total biological oxygen requirement was distributed to each basin as 22%, 55%, and 23%, respectively, based on past observations at the Nine Springs WWTP.

The additional oxygen requirement to maintain a specific level of DO in activated sludge was calculated using a mass balance approach for three aerobic basins with three separate DO concentrations. This mass balance is demonstrated in Equation S.3 [6]. Equation S.4 determined the total actual oxygen requirement for each of the three aerobic basins.

$$AOTR_{LIQUID} = \left[(Q_{IN} - Q_{RAS} \times DO) + (Q_{prev}(DO - DO_0)) \right] \times \frac{3785412 \text{ L/MG}}{453592 \text{ mg/lb}} \quad (\text{S.3})$$

$$AOTR_{TOTAL} = AOTR_{LIQUID} + R_o \quad (\text{S.4})$$

Where:

$AOTR_{LIQUID}$ = actual oxygen transfer rate to maintain DO concentration in the liquid,

(lb O₂/d)

Q_{IN} = influent flow rate with DO = 0 mg/L, million gallons per day (MGD),
assumed 40 MGD

Q_{RAS} = return activated flow rate, (MGD), assumed 0.5 Q_{IN}

Q_{prev} = flow rate from the previous basin, (MGD)

DO = concentration of DO in basin, (mg O₂/L)

DO₀ = concentration of DO in the first basin, (mg O₂/L)

AOTR_{TOTAL} = total actual oxygen requirement for maintaining a specific
concentration in the activated sludge and for the biological
requirement (lb O₂/d)

The field oxygen transfer efficiency was calculated in Equation S.5 with standard oxygen transfer efficiencies determined specifically for Nine Springs WWTP aeration basins from a previous study. Standard oxygen transfer efficiencies vary with the type of aeration device, basin geometry, degree of mixing, and the wastewater characteristics [6]. The standard oxygen transfer efficiencies in wastewater (α SOTE) determined for Nine Springs WWTP aeration basins were 0.10, 0.18, and 0.18 for the three aerobic zone passes, respectively. The field oxygen transfer efficiency assumed a water temperature of 20°C and saturated DO concentration of 10.6 mg O₂/L.

$$OTE_{20} = \alpha_{SOTE} \times \frac{C_{SAT} - DO}{C_{SAT}} \quad (S.5)$$

Where:

OTE₂₀ = field oxygen transfer efficiency at 20°C, unitless

α SOTE = standard oxygen transfer efficiency for wastewater, unitless

C_{SAT} = concentration of DO at saturation, (mg O₂/L), assumed 10.6 mg O₂/L

DO = concentration of DO in the aeration basin, (mg O₂/L)

The field oxygen transfer efficiency was applied to the AOTR_{TOTAL} for each aeration basin in Equation S.6. In addition, the resulting adjusted oxygen transfer rate was converted to an air flow rate in cubic feet per minute (CFM). The CFM for each aeration basin was summed to determine the total CFM required for the entire aerobic zone.

$$CFM = \frac{AOTR_{TOTAL}}{OTE_{20}} \times \frac{1}{60 \text{ min} \times \rho_{air} \times O_2} \quad (S.6)$$

Where:

CFM = required airflow rate, cubic feet per minute (ft³/min)

ρ_{air} = air density, pounds per cubic feet (lb/ft³), assumed 0.0765 lb/ft³

O₂ = fraction of oxygen in air, assumed 0.21

The total required airflow rate in CFM was converted to standard CFM (SCFM), using Equation S.7. This equation assumed 25% humidity, saturation pressure of 0.3631 psi, and 20°C air temperature. The SCFM airflow was incorporated into the calculation for the power requirement of the blowers in the mass flow rate of air (w) within Equation S.8 for U.S. Customary Units [6].

$$SCFM = \frac{CFM}{[P_{std}/(p_1 - P_{sat}\phi)] \times (T_{act}/T_{std})} \quad (S.7)$$

$$P_w = \left[\frac{wRT_{act}}{550ne} \left[\left(\frac{p_2}{p_1} \right)^{0.283} - 1 \right] \right] \times \frac{0.746kW}{hp} \times \frac{8760h}{yr} \quad (S.8)$$

Where:

SCFM = standard cubic feet per minute, (SCFM)

P_{std} = standard pressure, pounds per square inch (psi), 14.7 psi

P_{sat} = saturation pressure, pounds per square inch (psi), 0.3631 psi

p_1 = inlet pressure, (psi), 14.25 psi at the Nine Springs 877 ft elevation

Φ = blower air humidity, (%), assumed 25/100

T_{act} = actual inlet air temperature, Rankine ($^{\circ}\text{R}$), assumed 20 $^{\circ}\text{C}$ or 527.7 $^{\circ}\text{R}$

T_{std} = standard temperature, ($^{\circ}\text{R}$), assumed 520 $^{\circ}\text{R}$

P_w = power requirement for blowers, kilowatt hours per year (kWh/yr)

w = weight of flow of air, pounds per second (lb/s), $\text{SCFM} \times 0.0765 \text{ lb/ft}^3 \times 1\text{min}/60\text{s}$

R = engineering gas constant for air, 53.3 ft·lb/lb air· $^{\circ}\text{R}$

n = 0.283 for air

e = blower efficiency (usual range is 0.70-0.90), Nine Springs WWTP calculated average is 0.63

p_2 = absolute outlet pressure, (psi), assumed 22.6 psia based on 16 ft diffuser depth and 1 psi pressure loss

Supplementary References

1. Mobarry, B.K., et al., *Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria*. Appl Environ Microbiol, 1996. **62**(6): p. 2156-62. <https://www.ncbi.nlm.nih.gov/pubmed/8787412>
2. Amann, R., et al., *Combination of 16S rRNA-Targeted Oligonucleotide Probes with Flow Cytometry for Analyzing Mixed Microbial Populations*. Applied and Environmental Microbiology, 1990. **56**(6): p. 1919-1925. <http://aem.asm.org/content/56/6/1919.full.pdf>
3. Daims, H., et al., *The Domain-specific Probe EUB338 is Insufficient for the Detection of all Bacteria: Development and Evaluation of a more Comprehensive Probe Set*. Systematic and Applied Microbiology, 1999. **22**(3): p. 434-444. 10.1016/s0723-2020(99)80053-8
4. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis*. Nat Meth, 2012. **9**(7): p. 671-675. <http://dx.doi.org/10.1038/nmeth.2089>
5. Amann, R. and B.M. Fuchs, *Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques*. Nat Rev Microbiol, 2008. **6**(5): p. 339-48. 10.1038/nrmicro1888
6. Tchobanoglous, G.B., Franklin L.; Stensel, H. David, *Wastewater Engineering: Treatment and Reuse*. 4 ed, ed. M.E. Inc. 2003, New York, NY: McGraw-Hill. <http://www.mheducation.com/highered/product/wastewater-engineering-treatment-resource-recovery-metcalf-eddy-inc-tchobanoglous/M0073401188.html>

Chapter 3

Design and Assessment of Species-Level qPCR Primers Targeting Comammox

This chapter is published under the same title:

Beach, N.K. and D.R. Noguera, Design and Assessment of Species-Level qPCR Primers Targeting Comammox. *Frontiers in Microbiology*, 2019. 10(36). <https://doi.org/10.3389/fmicb.2019.00036>.

NKB and DRN developed the research plan and project goals. NKB designed the primers and performed laboratory work. Both NKB and DRN drafted the manuscript, tables and figures.

Abstract

Published PCR primers targeting the ammonia monooxygenase gene (*amoA*) were applied to samples from activated sludge systems operated with low dissolved oxygen (DO) to quantify total and clade-level *Nitrospira* that perform complete ammonium oxidation (comammox); however, we found these existing primers resulted in significant artifact-associated non-target amplification. This not only overestimated comammox *amoA* copies but also resulted in numerous false positive detections in the environmental samples tested, as confirmed by gel electrophoresis. Therefore, instead of attempting to quantify comammox diversity, we focused on accurately quantifying the candidate comammox species. We designed specific and sensitive primers targeting 3 candidate species: *Candidatus* (*Ca.*) *Nitrospira nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans*. The primers were tested with *amoA* templates of these candidate species and used to quantify comammox at the species level in low DO activated sludge systems. We found that comammox related to *Ca. N. nitrosa* were present and abundant in the majority of samples from low DO bioreactors and were not detected in samples from a high DO system. In addition, the greatest abundance of *Ca. N. nitrosa* was found in bioreactors operated with a long solids retention time. *Ca. N. inopinata* and *Ca. N. nitrificans* were only detected sporadically in these samples, indicating a minor role of these comammox in nitrification under low DO conditions.

Introduction

The oxidation of ammonium via nitrite to nitrate (i.e., nitrification) was historically considered a two-step process completed by phylogenetically distinct ammonia oxidizers and nitrite oxidizers. Although the complete oxidation of ammonium to nitrate by a single organism was theoretically possible, no microorganism with the ability to carry out both steps had been identified until late 2015 [1, 2]. These complete ammonium oxidizing (comammox) bacteria belong to the genus *Nitrospira*, which was known to contain only nitrite-oxidizing bacteria (NOB); therefore, an entire group of microorganisms with the ability to oxidize ammonia had been disguised for years [1, 2]. The discovery of comammox bacteria has redefined a key component of the global nitrogen cycle, initiating a renewed focus of recent nitrogen cycling research in environmental biotechnology [3-9].

Within the wastewater treatment industry, the application of nitrification under very low dissolved oxygen (DO) concentrations (below 0.2 mg O₂/L) is an exciting energy-saving approach to biological nutrient removal (BNR) [10-12]. Previous attempts to identify the key microorganisms responsible for nitrification in low DO bioreactors have been inconclusive, with the presence of known ammonia oxidizing bacteria (AOB) and ammonia oxidizing archaea (AOA) unable to explain observed nitrification rates [12]. Thus, the recent discovery of comammox in a variety of environments [1, 2] prompted us to investigate their presence in low DO BNR systems.

The gene encoding the alpha-subunit of ammonia monooxygenase (*amoA*) has been one of the most widely used markers for detection and quantification of AOB and AOA, as it facilitates functional analysis and reconstruction of phylogenetic relationships [13-

16]. The *amoA* of comammox is distinguishable from *amoA* sequences of AOB and AOA, and thus, it can be used to detect the presence of comammox in environmental samples [1, 2, 5, 17]. Recent descriptions of PCR primers targeting the *amoA* gene of commammox include a primer pair designed specifically for *Candidatus* (*Ca.*) *Nitrospira inopinata* [1], a primer pair designed specifically for a comammox-like clone within a freshwater aquaculture system [7], a primer collection that differentiates two broad clades of comammox within the *Nitrospira* genus [18], and a highly degenerate primer pair attempting to encompass all comammox within *Nitrospira* [8]. To date, three candidate comammox species have been described in the literature, namely *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans* [1, 2]. However, quantifying the contribution of these candidate species to the comammox community in environmental samples has not been possible because of the lack of specific qPCR primer sets. To overcome this limitation, we designed a set of highly specific non-degenerate primers for the independent detection of *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans*. We used these primers to evaluate the contribution of these species to the comammox population in samples from BNR plants operated at low DO conditions.

Methods

Sample collection, processing, and DNA extraction

Environmental samples used in this study originated from four low DO nitrifying bioreactors: a laboratory-scale sequencing batch reactor (L_SBR), a pilot-scale sequencing batch reactor (P_SBR), a pilot scale continuous flow (P_CF) reactor simulating a University of Cape Town configuration without nitrate recycle, and a full-

scale wastewater treatment plant (WWTP) from the Trinity River Authority (TRA) Central Region (Arlington, TX) (Table 3-1). For comparison, samples were also collected from the full-scale Nine Springs WWTP (NS) at the Madison Metropolitan Sewerage District (Madison, WI), which operates with typical high DO conditions (Table 3-1).

Table 3-1. Bioreactor sample characteristics

Bioreactor (Sample ID)	Location	Sample Dates ^a	Configuration at time of sampling	Size (gpd)	SRT (days)	Reference
Lab-scale sequencing batch reactor (L_SBR)	Madison, WI	3/7/2013 (1), 3/18/2013 (2), 4/17/2013 (3), 10/3/2014 (4)	Sequencing batch reactor (SBR) operated with dissolved oxygen (DO) < 0.2 mg O ₂ /L	0.5	80	Camejo, Owen [19]
Pilot-scale sequencing batch reactor (P_SBR)	Madison, WI	9/10/2015 (1), 11/10/2015 (2), 12/15/2015 (3), 4/25/2016 (4)	Sequencing batch reactor (SBR) operated with DO < 0.7 mg O ₂ /L	130	80	Camejo, Owen [19]
Continuous flow pilot plant (P_CF)	Madison, WI	9/9/2015 (1), 11/18/2015 (2), 12/30/2015 (3), 4/26/2016 (4)	Modified University of Capetown (UCT) operated with DO < 0.5 mg O ₂ /L	1200	10	Keene, Reusser [20]
Trinity River Authority Central Region Wastewater System Treatment Plant (TRA) ^b	Arlington, TX	8/9/2016 (1), 7/28/2017 ^b (2), 7/28/2017 (3)	Anaerobic Aerobic (AO) operated with three zones. Zone 1: DO < 0.6 mg O ₂ /L Zone 2: DO < 0.9 mg O ₂ /L Zone 3: DO < 1.4 mg O ₂ /L	123 M	10	This study
Madison Metropolitan Sewerage District Nine Springs Plant (NS)	Madison, WI	9/9/2015 (1), 11/18/2015 (2), 12/30/2015 (3), 4/26/2016 (4)	Modified University of Capetown (UCT) operated with DO > 2.0 mg O ₂ /L	40 M	10	Keene, Reusser [20]

a. Number in parentheses after each date is incorporated after the sample ID

b. TRA sample 2 and 3 were collected from different aeration basins

Grab samples with varying volumes were collected from every reactor and were either centrifuged to form a biomass pellet, with supernatant discarded, and the pellet frozen at -80 °C (P_SBR, P_CF, TRA, and NS) or were saved in a glycerol mixture prior to placement in a -80 °C freezer (L_SBR). Biomass samples were stored at -80 °C until DNA extraction. All DNA was extracted using DNeasy® PowerSoil® DNA Isolation Kit (Qiagen, Hilden, Germany) following the manufacturer's directions. DNA was quantified with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) and the purity ratio, or ratio of absorbance at 260 nm and 280 nm, was determined with a NanoDrop

spectrophotometer (Thermo Fisher Scientific, Waltham, MA). DNA samples were stored at -20 °C until further processing.

Design of primers to detect comammox ammonia monooxygenase gene amoA

A collection of full-length *amoA* and full-length particulate methane monooxygenase (*pmoA*) gene sequences, obtained from the National Center for Biotechnology Information (NCBI) GenBank database [21], were used for primer design. The full-length sequences included in the design are indicated by bold text in the phylogenetic tree in Figure 3-3, with the corresponding accession number in parenthesis. All *amoA* and *pmoA* sequences were aligned using the 'AlignSeqs' command in the DECIPHER "R" package [22, 23]. This aligned database was then submitted to DECIPHER's Design Primers web tool [24]. Sequences corresponding to *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans* were individually selected as target groups for primer design, which used the following parameters: primer length ranging from 17 - 26 nucleotides with up to 1 permutation, PCR product amplicon length of 100 - 450 bp, 100% target group coverage, and without the Taq 3'-end Model option. The primer design tool also used the same reaction conditions in all cases: [Na⁺] 70 mM, [Mg²⁺] 3 mM, [dNTPs] 0.8 mM, annealing temperature (T_a) of 64 °C, [primers] 400 nM. Amplification products were verified by agarose (2%) gel electrophoresis with GelRed™ Nucleic Acid Gel Stain (Biotium, Freemont, CA).

Quantitative real-time polymerase chain reaction (qPCR)

Quantification of total 16S rRNA genes, total comammox *amoA* genes, clades A and B comammox *amoA*, as well as *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans*

amoA in each DNA sample was carried out by qPCR. All qPCR assays were performed on a Roche LightCycler® 480 high-throughput real-time PCR system using white LightCycler® 480 multiwell plates and the associated LightCycler® 480 sealing foils (Roche Molecular Systems, Inc., Pleasanton, CA, USA). All environmental DNA samples were diluted to 10 ng/μL. Although not specifically studied here, the applied sample dilutions helped reduce any potential problems with PCR inhibition. All qPCR assays were prepared in Bio-Rad 2x iQ™ SYBR® Green Supermix (Bio-Rad, Hercules, CA, USA), containing 50 U/mL iTaq DNA polymerase, 1.6 mM dNTPs, 100 mM KCl, 40 mM Tris-HCl, 6 mM MgCl₂, 20 nM fluorescein, and stabilizers (10 μL per reaction). Triplicate reactions were prepared for each sample.

Amplification of total comammox *amoA* was performed according to Fowler, Palomo [8], with each reaction containing Bio-Rad 2x iQ™ SYBR® Green Supermix (10 μL), nuclease free water (4.4 μL), environmental DNA or standard (4 μL), and PCR primers (0.8 μL of 12.5 μM each primer) added for a final volume of 20 μL per reaction. The thermal cycling protocol for the total comammox primers was as follows: initial denaturation step at 94 °C for 5 min, followed by 40 cycles of initial denaturation at 94 °C for 30 s, annealing at 48 °C for 30 s, and extension at 72 °C for 1 min.

Amplification of clade A and clade B comammox *amoA* was performed using the equimolar primer mixtures according to Pjevac, Schauburger [18] (Table 3-2). Four equimolar primer mixtures were created for clade A and clade B forward and reverse primers, by combining 62.5 μL of each of the six individual primer stocks (100 mM) together with 125 μL nuclease free water to a final working volume of 500 μL and final concentration of 12.5 μM each primer. Finally, reactions were prepared in Bio-Rad 2x

iQTM SYBR[®] Green Supermix (10 µL), nuclease free water (4.4 µL), environmental DNA or standard (4 µL), and equimolar PCR primer mixture (0.8 µL of 12.5 µM each primer) were added to each reaction for a final volume of 20 µL per reaction. The thermal cycling protocol for the clade-level primers was as follows: initial denaturation step at 95 °C for 3 min, followed by 45 cycles of initial denaturation at 95 °C for 30 s, annealing at 52 °C for 45 s, and extension at 72 °C for 1 min.

For the novel *amoA* primers designed in this study (Table 3-2), nuclease free water (4.4 µL), environmental DNA or standard (4 µL), and PCR primers (0.8 µL of 10 µM each primer) were added to each reaction for a final volume of 20 µL per reaction. The thermal cycling protocol for qPCR using the novel *amoA* primers designed in this study (Table 3-2) was as follows: initial denaturation step at 95 °C for 10 min, followed by 45 cycles of initial denaturation at 95 °C for 10 s, annealing at 64 °C for 30 s, and extension at 72 °C for 30 s. Fluorescence was measured at 72 °C for amplicon quantification. After amplification, an amplicon melting curve was recorded in 0.25 °C steps between 65 and 97°C. Melting peaks were obtained by plotting the negative first derivative of fluorescence against temperature. Although 30 cycles is typically sufficient for quantification of targets in qPCR, the thermal cycling was extended to evaluate potential non-specific amplification with the newly designed primers [24]. Finally, amplification of total 16S rRNA was performed using the 16S rRNA-targeted primer pair 341f/785r [25] according to Thijs, Op De Beeck [26] (Table 3-2).

Table 3-2. Primers used for qPCR

Target gene	Primer name		Forward primer (5' - 3')	Reverse primer (5' - 3')	T _a (°C)	Reference
<i>Ca. Nitrospira nitrosa amoA^a</i>	Nitrosa 469F/812R	amoA-	GCGATTCTGTTTTATCCCAGCAAC	CCGTGTGCTAACGTGGCG		
<i>Ca. Nitrospira inopinata amoA^a</i>	Inopinata 410F/815R	amoA-	TCACCTTGTTGCTAACTAGAACTGG	TCCGCGTGAGCCAATGT	64	This study
<i>Ca. Nitrospira nitrificans amoA^a</i>	Nitrificans 463F/836R	amoA-	ATGTTTCGCGGCACTGTT	CCAGAAAGTTTAGCTTTGTCGCCT		
<i>Comammox Nitrospira clade A amoA^b</i>	comaA-244f_a/659r_a		TACAACTGGGTGAACTA	AGATCATGGTGCTATG	52	Pjevac et al. (2017)
	comaA-244f_b/659r_b		TATAACTGGGTGAACTA	AAATCATGGTGCTATG		
	comaA-244f_c/659r_c		TACAATTGGGTGAACTA	AGATCATGGTGCTGTG		
	comaA-244f_d/659r_d		TACAACTGGGTCAACTA	AAATCATGGTGCTGTG		
	comaA-244f_e/659r_e		TACAACTGGGTCAATTA	AGATCATCGTGCTGTG		
	comaA-244f_f/659r_f		TATAACTGGGTCAATTA	AAATCATCGTGCTGTG		
<i>Comammox Nitrospira clade B amoA^b</i>	comaB-244f_a/659r_a		TAYTTCTGGACGTTCTA	ARATCCAGACGGTGTG	52	Pjevac et al. (2017)
	comaB-244f_b/659r_b		TAYTTCTGGACATTCTA	ARATCCAAACGGTGTG		
	comaB-244f_c/659r_c		TACTTCTGGACTTTCTA	ARATCCAGACAGTGTG		
	comaB-244f_d/659r_d		TAYTTCTGGACGTTTTA	ARATCCAAACAGTGTG		
	comaB-244f_e/659r_e		TAYTTCTGGACATTTTA	AGATCCAGACTGTGTG		
	comaB-244f_f/659r_f		TACTTCTGGACCTTCTA	AGATCCAAACAGTGTG		
Total comammox <i>Nitrospira amoA^c</i>	Ntsp-amoA 162F/359R		GGATTCTGGNTSGATTGGA	WAGTTNGACCACCASTACCA	48	Fowler et al. (2017)
Total 16S rRNA ^d	16S-341f/785r		CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	53	Klindworth et al. (2013); Thijs et al. (2017)

Standard preparation for qPCR

Full-length *amoA* from *Candidatus Nitrospira nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans* were used as standards for the new species-specific comammox assay as well as the total comammox and clade A comammox assays. Full-length *amoA* from *Nitrospira sp. CG24_E* [27] was used as the standard for the clade B comammox assay. These individual *amoA* standards were generated from synthetic gene plasmid cloning vectors (Integrated DNA Technologies, Inc., Coralville, IA) transformed into One Shot™ TOP10 Competent *E. coli* (Life Technologies, Carlsbad, CA) with the TOPO-TA cloning kit (Invitrogen, Karlsruhe, Germany) which uses the pCR™4-TOPO® TA vector. The cloned plasmids were subjected to amplification with M13 primers and product sizes verified with gel electrophoresis. The M13-PCR products were purified using the Qiagen PCR Purification Kit (Qiagen, Hilden, Germany) and quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA). The purity ratio of the standard was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Standard gene copy number was calculated from the DNA concentration. The standards were diluted in a 10-fold series over seven orders of magnitude (approximately 10^1 to 10^7 copies) and were included in triplicate with each set of samples. A 95% confidence interval and standard deviation was calculated for the fractional PCR cycle used for quantification (i.e. quantification cycle or C_q) for each set of standard replicates. A regression line fit to the C_q and the known concentration of each standard was used to generate a standard calibration curve for each assay. Next, the standard replicate statistics and assay-specific standard calibration curve were used to approximate a limit

of detection (LOD) and limit of quantification (LOQ) according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE Guidelines) described by Bustin, Benes [28] and equations adapted from Armbruster and Pry [29]. Median LOD, median LOQ, and corresponding 95% confidence intervals are reported for each assay (Table 3-3). External standard curve amplification efficiencies were calculated as $E = (10^{\frac{-1}{m}} - 1) \times 100$, where m represents the slope of the standard curve.

Gradient PCR

Gradient PCR was performed with the total comammox [8] and clade-level comammox [18] *amoA* primers with environmental DNA samples diluted to three concentrations to determine if an optimal annealing temperature and DNA concentration exists to eliminate or at least minimize the observed non-specific amplification. Gradient PCR was performed on a Mastercycler® nexus gradient PCR cycler (Eppendorf North America, Hauppauge, NY) with Eppendorf twin.tec® 96-well LoBind PCR plates and Eppendorf PCR sealing foils. Reaction mixtures and thermal cycling protocols were identical to the respective qPCR assay, with the exception that annealing temperatures were set to vary across each PCR plate. For the total comammox primers, annealing temperatures were set in small increments between 47.8 °C and 56 °C. For the clade-level comammox primers, annealing temperatures were set in small increments between 51.8 °C and 60 °C. Two samples were selected for each assay and were selected based on the presence of both the target and non-specific

Table 3-3. qPCR performance with three distinct comammox *amoA* standard templates. Average parameters are presented with standard deviation. PCR efficiency is reported for the linear range of the standard calibration curve.

	Primer set	Amplicon size (bp)	Amplicon GC content (%)	Average T_m (°C)	Log ₁₀ linear range, log(SQ)*	Average PCR efficiency (%)	Median LOD** (copies/ng DNA)	Median LOQ** (copies/ng DNA)
Comammox <i>Nitrospira</i> clade A	Standard template: <i>N. nitrosa</i> <i>amoA</i>							
	Nitrosa amoA-469F/812R ^a	344	57	88.8 ± 0.08	1 - 7	96 ± 3.0	M= 11, 95% CI [9.3, 13]	M= 13, 95% CI [10, 17]
	Ntsp-amoA 162F/369R ^b	198	54	86.1 ± 0.08	3 - 7	91 ± 2.6	M= 20, 95% CI [8.6, 48]	M= 28, 95% CI [11, 72]
	comaA-244f/659r ^c	415	58	88.9 ± 0.09	2 - 7	95 ± 2.7	M= 38, 95% CI [13, 110]	M= 40, 95% CI [14, 120]
	Standard template: <i>N. inopinata</i> <i>amoA</i>							
	Inopinata amoA-410F/815R ^a	406	49	85.6 ± 0.05	1 - 7	90 ± 3.0	M= 7.0, 95% CI [4.0, 12]	M= 14, 95% CI [7.0, 29]
	Ntsp-amoA 162F/369R ^b	198	49	84.4 ± 0.60	3 - 7	82 ± 3.4	M= 20, 95% CI [8.6, 48]	M= 28, 95% CI [11, 72]
	comaA-244f/659r ^c	415	50	86.0 ± 0.10	3 - 7	86 ± 2.5	M= 38, 95% CI [13, 110]	M= 40, 95% CI [14, 120]
	Standard template: <i>N. nitrificans</i> <i>amoA</i>							
	Nitrificans amoA-463F/836R ^a	374	52	86.7 ± 0.10	1 - 7	92 ± 2.4	M= 1.3, 95% CI [0.9, 2.0]	M= 3.8, 95% CI [1.3, 11]
	Ntsp-amoA 162F/369R ^b	198	57	87.3 ± 0.27	2 - 7	96 ± 1.8	M= 20, 95% CI [8.6, 48]	M= 28, 95% CI [11, 72]
	comaA-244f/659r ^c	415	56	88.2 ± 0.07	2 - 7	93 ± 2.9	M= 38, 95% CI [13, 110]	M= 40, 95% CI [14, 120]

* SQ = starting quantity in copies per reaction

** M = median, CI = confidence interval [upper limit, lower limit] ;

a. This study

b. Fowler et al., 2017

c. Pjevac et al., 2017

amplifications after qPCR. The total comammox gradient assay included DNA from P_SBR-4 and NS-4. The clade A and clade B comammox gradient assays included DNA from P_CF-4 and NS-4. All environmental DNA samples were diluted to 40 ng DNA/reaction, 20 ng DNA/reaction, and 1 ng DNA/reaction. Amplification products were then visualized with agarose (2%) gel electrophoresis with GelRed™ Nucleic Acid Gel Stain (Biotium, Freemont, CA).

Phylogenetic tree construction

The database of aligned *amoA* and *pmoA* sequences was expanded by including additional comammox *amoA* sequences obtained from metagenomes available on the JGI Integrated Microbial Genomes (IMG) database [30] and available comammox *Nitrospira* bins extracted with mmgenome (<http://madsalbertsen.github.io/mmgenome/>) [31]. The expanded database was utilized for *in-silico* primer analyses and to construct a consensus phylogenetic tree in Geneious v. 9.1.2 [32] using the Neighbor-Joining method on a Tamura-Nei genetic distance model with bootstrap resampling and 20,000 trials.

Average nucleotide identity calculation

To determine if comammox *Nitrospira* draft genomes belonged to the same species, we performed pairwise analyses of genome average nucleotide identity (ANI). Initially, open reading frames (ORFs) within each draft genome was predicted with Prodigal [33]. Subsequently, the JSpecies Web Server [34] was used to calculate ANI based on BLAST+ [35] from percent of aligned genome ORFs.

Results

Total comammox and clade-level comammox detection

For broad comammox *amoA* quantification, three published primer sets were compared using standards and nineteen environmental samples originating from five bioreactors (Table 3-1). The Ntsp-*amoA* 162F/359R primer set [8], designed to target total comammox *Nitrospira amoA* (Table 3-2), amplifies a 198 bp fragment at primer binding regions between positions 162-182 and 339-359 bp (Figure 3-1). The comaA-244f/659R and comaB-244f/659r primer sets [18], designed to differentiate between clades A and B of comammox (Table 3-2), amplify a 415 bp fragment at *amoA* primer binding regions between positions 244-261 and 643-659 bp (Figure 3-1). The standard curves for these assays were linear over a minimum of five orders of magnitude and had high coefficients of determination ($R^2 > 0.990$); although linear range, amplification efficiency, LOD and LOQ varied depending on the comammox *amoA* standard used (Table 3-3; Table S3-1). The results for the total comammox [8] and clade-level comammox [18] assays are also summarized in decision matrices that were used to consistently evaluate qPCR results between assays (Supplementary Tables S3-3 to S3-5).

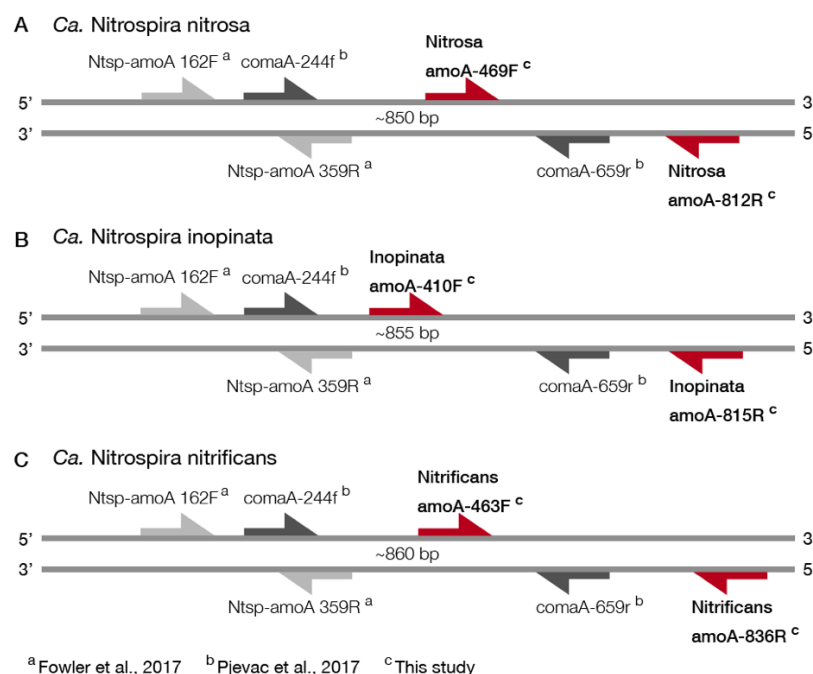
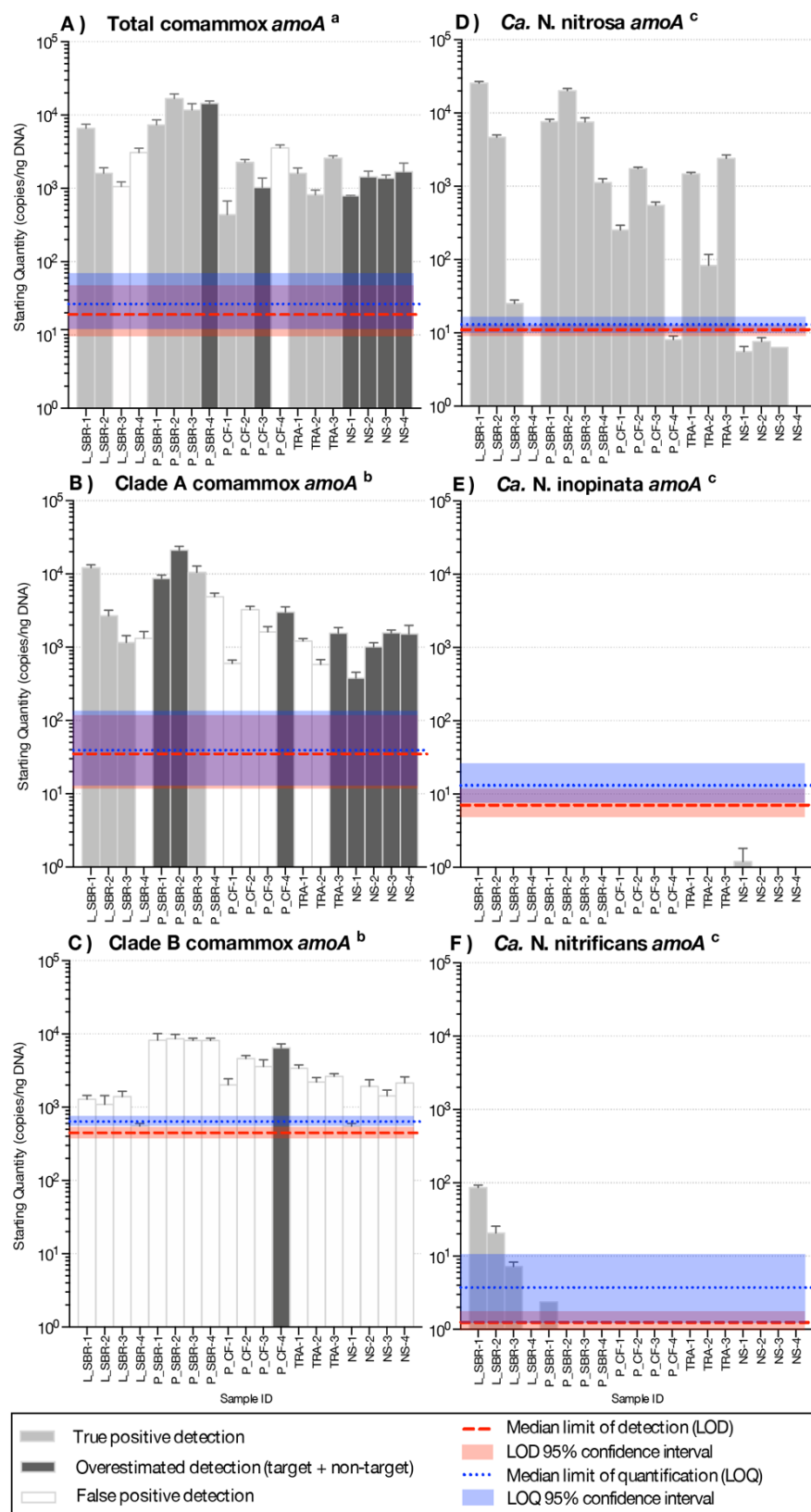


Figure 3-1. Visual representation of primer amplification regions with A) *Ca. N. nitrosa* *amoA*, B) *Ca. N. inopinata* *amoA*, and C) *Ca. N. nitrificans* *amoA*. The new primers, labeled with red arrows, were designed using full-length *amoA* sequences and novel primer binding region was discovered near the end of each *amoA* gene (approximately between 400 and 800 bp). Base pair locations are indicated in the primer names. Most publicly available published *amoA* sequences are partial length and are located between approximately 244 bp and 659 bp. Amplification with clade A-specific equimolar primer mixtures occur in this region, while amplification with the total comammox primers occur between 162 bp and 359 bp.

In the total comammox assay, melting curves for the standards were narrow and unimodal above $\sim 10^1$ - 10^2 copies (Supplementary Figure S3-1 (A-C)). However, melting curves from the environmental samples (Figure S3-1 (D-H)) and the agarose gel of qPCR products (Figure S3-2) showed non-specific amplifications resulting in three false positive detections (amplification did not correspond to expected product size) and overestimation of comammox abundance (multiple products seen, including one with the expected product size) in six samples (Figure 3-2A; Table S3-3).



^a Fowler et al., 2017, ^b Pjevac et al., 2017, ^c This study

Figure 3-2. Abundance of *amoA* genes (gene copies/ng DNA) on a log-scale from comammox *Nitrospira* using the (A) total comammox primers, (B) clade A primers (C) clade B primers, (D) the newly designed *Ca. N. nitrosa* specific primers (E) *Ca. N. inopinata* specific primers and (F) *Ca. N. nitrificans* specific primers from various time series samples originating from five bioreactors. LS, lab-scale sequencing batch reactor; P_SBR, pilot-scale sequencing batch reactor; P_CF, continuous flow pilot plant; TRA, Trinity River WWTP; NS, Nine Springs WWTP. Error bars show the standard deviation of the triplicate samples. The color of each column indicates whether there was a true positive detection, an overestimated detection, or a false positive detection. The median limit of detection and quantification for each assay are represented by horizontal red and blue dotted lines, respectively. The shaded regions surrounding these limits represent the upper and lower bounds of the 95% confidence interval.

In the clade A comammox assay, all standards with approximately 10^2 *amoA* copies or less presented two melting peaks, the expected target peak at approximately 88.9 °C, 86.0 °C, and 88.2 °C for *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA*, respectively, plus a sizeable off-target peak at approximately 70 °C (Figure S3-3 (A-C)). Higher copy numbers produced narrow, unimodal melting peaks at the expected T_m . The agarose gel performed with the qPCR products showed a single band for the standards; however, eight of the nineteen environmental samples contained multiple non-specific amplified products including a potential target amplicon (Figure S3-3 (D-H); Figure S3-4) leading to overestimated gene abundance (Figure 3-2B; Table S3-4). Amplification of non-specific products in seven additional samples produced artifact-associated fluorescence that was well over the LOQ and translated into false positive predictions of the same order of magnitude as true positive detections (Figure 3-2B).

Clade B standards with approximately 10^3 *amoA* copies or less also contained two melting peaks, an off-target peak at approximately 72 °C and the expected target peak at approximately 89 °C (Figure S3-5 A). In the clade B assay, fifteen samples had a target melting peak at the expected T_m (Figure S3-5 (B-F)), however, the agarose gel revealed that these samples had abundant non-specific products that were approximately 300 bp long— smaller than the expected 415 bp target size (Figure S3-6). Only one sample (P_CF-4) appeared to contain an amplicon of the expected size (Figure S3-6), but also contained several other non-specific amplicons. Thus, none of the samples tested were confirmed to be a true positive for presence of clade B comammox (Table S3-5).

The biggest challenge with the non-specific amplification observed with the total comammox and clade-level comammox primers was the significant artifact-associated fluorescence resulting in C_q values within the same range as samples with abundant comammox (see Supplementary Tables S3-3, S3-4, and S3-5), which did not allow us to eliminate samples based on the LOD or LOQ. Two factors that could have contributed to the non-specific amplifications observed with the broad-detection primers are the low design annealing temperatures (48 °C for the total comammox primers and 52 °C for the clade A and clade B comammox primers) and the total DNA concentration we used in each assay. Therefore, we performed gradient PCR experiments to evaluate whether adjustments to the primer annealing temperature or sample DNA concentration would improve the specificity of the total comammox primers [8] and the clade A and clade B comammox primer sets [18]. In our specific application, the gradient PCR experiments show that small adjustments to annealing temperature or sample dilutions did not make a meaningful difference in the non-specific amplifications. Non-specific products remained in the samples tested at higher annealing temperatures and at all three DNA concentrations tested (40 ng/reaction, 20 ng/reaction and 1 ng/reaction) with all three published primer sets. In the gradient PCR with the total comammox primers, non-specific products are observed at approximately 300 bp and 600 bp in the P_SBR-4 sample and at approximately 150 bp, 350 bp, 600 bp and 700 bp in the NS-4 sample, and do not appear to be dependent on the total DNA concentrations tested in the reaction (Figure S3-7). In the gradient PCR for the clade A comammox primers, non-specific products are observed at approximately 150 bp and 300 bp in the P_CF-4 sample and at approximately 250 bp

and 300 bp in the NS-4 sample, and the amplification of these non-specific products do not appear to be dependent on the total DNA concentrations tested in the reaction (Figure S3-8). In the gradient PCR for the clade B comammox primers (Figure S3-9), multiple non-specific products were observed, with the most prominent non-specific product consistent with previous observations at approximately 300 bp.

Another challenge we faced when applying the broad-detection comammox primers to our samples was the inability to rely on melting curve analysis alone to differentiate between true positive detections, false positive detections, and overestimated detections. To successfully complete melting curve analysis post qPCR, the melting characteristics of an amplified standard (characterized by the both the shape of the melting peak and the temperature where the peak occurs or T_m) is compared to the melting peaks that appear in the environmental samples since the T_m considered a unique feature of a particular DNA fragment [36, 37]. However, this type of analysis can be difficult with primer sets that target a broad number of sequences, especially when the primers are designed to amplify a region with highly variable GC content. In this study, we found that the total comammox primers [8] amplify a region of *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA* with GC content at 54%, 49%, and 57%, respectively. This produced melting peaks for *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA* at a T_m of 86.1 °C, 84.4 °C, and 87.3 °C, respectively. We also found that the clade A comammox primers [18] amplify a region of *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA* with GC content at 58%, 50%, and 56%, respectively. This, in turn, produced melting peaks for *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA* at a T_m of 88.9 °C, 86.0 °C, and 88.2 °C, respectively. While

it is possible to determine a T_m range based on in silico analysis of the known comammox diversity, this does not allow us to eliminate non-specific products from the analysis since some may contain similar melting characteristics as the target. For example, the melting peaks for P_CF-4 with the total comammox primers (Figure S3-1 F) fall within the T_m range set by the standards used in this study, but only contained a non-specific product between 400 and 500 bp long when verified with agarose gel electrophoresis (Figure S3-2). Since the T_m is also influenced by the reaction conditions, one would need to include a much greater number of standards in situ to represent a greater variety of known comammox species in order to truly rely on melting curve analysis alone with these broad-detection assays. Since more species are continually discovered, relying on melting curve analysis alone to confirm positive detections will likely be difficult without an array of comammox *amoA* standards and will require confirmation with an agarose gel electrophoresis of amplified qPCR products.

Overall, the presence, frequency, and influence of unspecific amplification in the total comammox [8] assay and in the clade A and clade B [18] comammox assays were significant and did not allow for accurate detection and quantification of comammox *Nitrospira* in the specific environmental samples tested in this study (Figure 3-2 (A-C)), prompting us to design other options for specific comammox detection.

Design of species-specific primer sets

Since the broad range comammox primers did not provide satisfactory results with our environmental samples, we opted to focus our evaluations on the presence of specific comammox species. With this objective in mind, to detect and quantify comammox

belonging to the *Candidatus Nitrospira* species currently described in the literature (*Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans*) we designed qPCR primer sets specifically targeting each of these species. For this design, we use a dataset of 27 full-length *amoA* gene sequences from AOB, AOA, and comammox, and 15 full-length *pmoA* gene sequences from methanotrophs. We limited the database to full-length *amoA* and *pmoA* sequences in order to allow for discovery of primer-binding regions outside of the fragments amplified by conventional *amoA* primer sets [13, 38]. We used the Design Primers option in DECIPHER [24] with each one of the candidate species as the target group and other *amoA/pmoA* sequences entered as closely related groups that should not be amplified. In addition, for all designs, the annealing temperature and PCR conditions were fixed (see Materials and Methods) so that all three species-level primer sets could be simultaneously used in a single thermocycler run. For these species-specific primers, we take advantage of a higher annealing temperature ($T_a = 64\text{ }^{\circ}\text{C}$) than what was used for the total comammox ($T_a = 48\text{ }^{\circ}\text{C}$) and clade-level comammox ($T_a = 52\text{ }^{\circ}\text{C}$) primers, which is a strategy that also helps improve primer specificity.

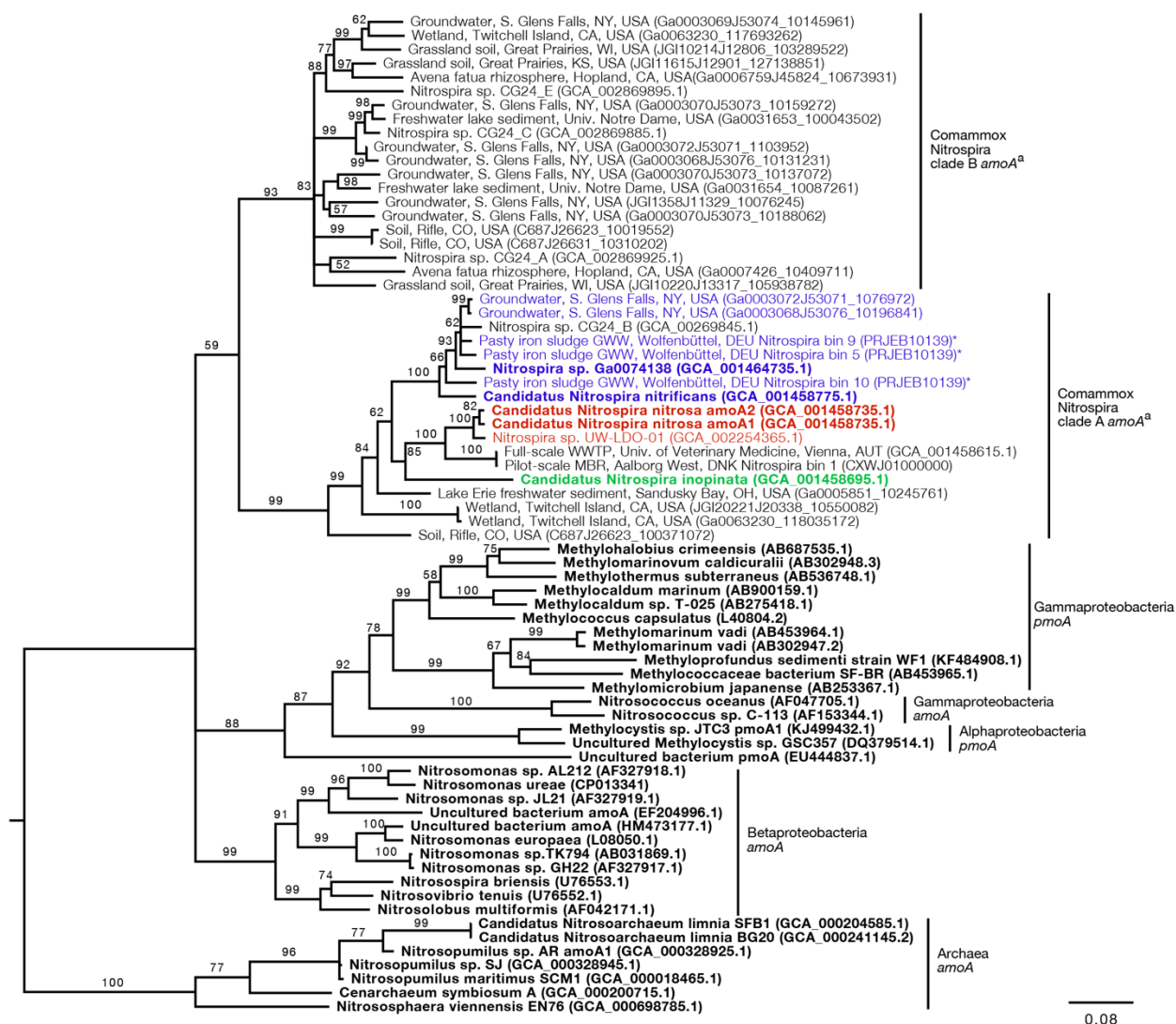
For *Ca. N. nitrosa*, the two *amoA* copies identified in the *Ca. N. nitrosa* genome [2] were used as the target group. The design algorithm identified primer binding regions at positions 469-493 and 794-812, predicting amplification of a 344 bp fragment (Figure 3-1). The target regions for each primer had 5 or greater mismatches to all other sequences in the design database, predicting a 100% specificity to the target group.

A single sequence [1] served as the target group for the design of *Ca. N. inopinata* primers, resulting in primer binding sites at 410-436 and 798-815, for a 406 bp fragment amplification. The target region for each primer had 5 or greater mismatches to other

amoA and *pmoA* sequences used as non-targets in the design database, also predicting 100% specificity.

For *Ca. N. nitrificans*, two sequences were used as targets; the one in the *Ca. N. nitrificans* genome [2] and the one from *Nitrospira* sp. Ga0074138 [5], which had 91.4% sequence identity to the *Ca. N. nitrificans* sequence. The primers amplified a 374 bp fragment, between positions 463 and 836. The predicted specificity was again 100%, with 5 or greater mismatches per primer to other sequences used in the design database.

Since the sequence dataset used for design was small, we searched for additional *amoA* sequences that would contain the target regions of the new primer sets. However, this analysis was limited because most of the published *amoA* sequences contain the target side of the forward primers, but not the site targeted by the reverse primer (Figure 3-1). In this search, we found 33 nearly full-length sequences that not only clustered with the comammox sequences but were also long enough to contain the target regions of the species-specific primers (Figure 3-3). This set of nearly full-length *amoA* sequences



^aComammox Nitrospira clade A and clade B amoA accession numbers beginning with 'Ga', 'JGI', or 'C687' are JGI IMG scaffold IDs

Figure 3-3. Neighbor-joining consensus tree generated from an alignment of full-length and near full-length *amoA* and *pmoA* gene sequences (> 600bp), rooting with archaea *amoA*. Bootstrap values, shown at the nodes where the value was greater than 50, are based on 10,000 trials. Bold sequence names were included in the new primer design. Blue, red, and green text are amplified by the *Ca. N. nitrificans*, *Ca. N. nitrosa*, and *Ca. N. inopinata* primer sets, respectively. Asterisks indicate that the gene is amplified but with a reduced efficiency due to base pair mismatches. The scale bar indicates the number of nucleotide substitutions per site. Accession numbers are presented after the sequence names. Acronyms were used for groundwater well (GWW), wastewater treatment plant (WWTP), and membrane bioreactor (MBR).

included *amoA* sequences from comammox metagenomes that have been recently described in the literature [1, 2, 4, 5, 17, 18].

An *in-silico* primer-target analysis (Figure 3-4) shows that *N. sp.* UW-LDO-01 [17], originally described as a strain of *Ca. N. nitrosa*, has perfect matches to the newly designed primers targeting this species. Two *amoA* sequences (Full scale WWTP, Univ. of Veterinary Medicine, Vienna, AUT, GCA_001458615.1; Pilot-Scale MBR, Aalborg West, DNK Nitrospira bin 1, CXWJ01000000) clustered together and near the *Ca. N. nitrosa* cluster (Figure 3-3). The average nucleotide identity (ANI) of their genomes compared to *Ca. N. nitrosa* is 84.5% and 85%, respectively (Figure 3-5), lower than the typical cutoff for species definition (ANI > 94%) [39]. In agreement, we predict that their *amoA* will not be amplified by the nitrosa-specific primer set due to multiple mismatches with forward and reverse primers (Figure 3-4). The *amoA* sequence of *Nitrospira sp.* CG24_B [4] clustered with *Ca. N. nitrificans* (Figure 3-3). However, the ANI of these two genomes (86.3%; Figure 3-5) is below the typical cutoff for species definition and the *in-silico* analysis predicts that *Nitrospira sp.* CG24_B *amoA* will not be amplified with the nitrificans-specific primers (Figure 3-4). Three additional *amoA* sequences clustering with *Ca. N. nitrificans* include those from the draft genomes of ‘Pasty iron sludge GWW N. bin 5, N. bin 9, and N. bin 10 (PRJEB10139)’ [1] (Figure 3-3), which contain fewer mismatches (Figure 3-4) and are predicted to partially amplify (below 56% efficiency) with the *Ca. N. nitrificans* primers despite ANI below the species cutoff (Figure 3-5).

Primer Set and Targets	Forward Target Sequence																		Reverse Target Sequence																		Predicted PCR Efficiency (%)										
Nitrificans amoA-463F/836R	5'	A	T	G	T	T	C	G	C	G	G	C	A	C	T	G	T	T	3'	5'	C	C	A	G	A	A	A	G	T	T	T	A	G	C	T	T	T	G	T	C	G	C	C	T	3'	Forward / Reverse / Combined	
Groundwater, S. Glens Falls, NY, USA (Ga0003072J53071_1076972)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	96.3 / 99.1 / 97.7	
Groundwater, S. Glens Falls, NY, USA (Ga0003068J53076_10196841)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	96.3 / 99.1 / 97.7		
Nitrospira sp. CG24_B (GCA_00269845.1)	3'	.	.	.	.	.	.	.	G	.	.	C	A	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	5'	0.00 / 32.8 / 0.00			
Pasty iron sludge GWW, Wolfenbüttel, DEU Nitrospira bin 9 (PRJEB10139)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	5'	96.3 / 32.8 / 56.2		
Pasty iron sludge GWW, Wolfenbüttel, DEU Nitrospira bin 5 (PRJEB10139)	3'	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	1.14 / 99.1 / 10.6			
Nitrospira sp. Ga0074138 (GCA_001464735.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	96.3 / 99.1 / 97.7			
Pasty iron sludge GWW, Wolfenbüttel, DEU Nitrospira bin 10 (PRJEB10139)	3'	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	24.1 / 99.1 / 48.9			
Candidatus Nitrospira nitrificans (GCA_001458775.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	96.3 / 99.1 / 97.7			
Candidatus Nitrospira nitrosa amoA 1&2 (GCA_001458735.1)	3'	.	.	.	A	.	.	.	A	T	T	.	.	.	.	.	.	5'	3'	T	T	C	C	C	C	C	C	A	A	C	T	T	T	.	G	C	C	.	T	A	T	.	G	5'	0.00 / 0.00 / 0.00		
Candidatus Nitrospira inopinata (GCA_001458695.1)	3'	.	.	.	A	.	.	.	A	T	T	T	.	.	.	.	.	5'	3'	.	.	C	C	C	.	G	A	G	A	G	T	T	T	G	G	C	C	.	T	A	T	.	C	5'	0.00 / 0.00 / 0.00		
Nitrospira sp. CG24_A (GCA_002869925.1)	3'	G	C	.	.	.	.	.	.	A	T	G	G	.	A	.	.	5'	3'	.	.	.	T	C	C	G	.	C	G	.	G	C	.	A	.	G	T	C	T	C	T	G	.	5'	0.00 / 0.00 / 0.00		
Nitrosa amoA-469F/812R	5'	G	C	G	A	T	T	C	T	G	T	T	T	T	A	T	C	C	C	A	G	C	A	A	C	3'	5'	C	C	G	T	G	T	G	C	T	A	A	C	G	T	G	G	C	G	3'	
Candidatus Nitrospira nitrosa amoA 1&2 (GCA_001458735.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	99.0 / 99.1 / 99.0		
Nitrospira sp. UW-LDO-01 (GCA_002254365.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	99.0 / 99.1 / 99.0			
Full-scale WWTP, Univ. of Vet. Med., Vienna, AUT (GCA_001458615.1)	3'	.	.	T	.	.	.	.	T	.	.	.	.	.	.	T	.	5'	3'	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	0.05 / 75.1 / 0.00			
Pilot-scale MBR, Aalborg West, DNK Nitrospira bin 1 (CXXWJ01000000)	3'	.	.	T	.	.	.	.	T	.	.	.	.	.	.	T	.	5'	3'	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	0.05 / 75.1 / 0.00				
Candidatus Nitrospira nitrificans (GCA_001458775.1)	3'	.	.	G	C	A	.	.	.	.	C	.	.	.	.	G	.	5'	3'	T	G	T	G	.	C	C	.	C	.	.	G	.	.	.	.	A	5'	0.00 / 0.00 / 0.00									
Candidatus Nitrospira inopinata (GCA_001458695.1)	3'	.	.	.	.	T	.	.	.	C	.	.	.	G	.	.	T	5'	3'	G	.	.	.	A	.	C	.	.	T	.	.	T	.	.	A	5'	0.01 / 0.00 / 0.00										
Nitrospira sp. CG24_A (GCA_002869925.1)	3'	.	.	.	.	G	G	.	A	.	.	A	.	.	T	C	A	.	T	5'	3'	G	.	A	A	A	C	T	.	A	T	C	T	.	.	C	.	A	5'	0.00 / 0.00 / 0.00							
Inopinata amoA-410F/815R	5'	T	C	A	C	C	T	T	G	T	T	G	C	T	A	A	C	T	A	G	A	A	A	C	T	G	G	3'	5'	T	C	C	G	C	G	T	G	A	G	C	C	A	A	T	G	T	3'
Candidatus Nitrospira inopinata (GCA_001458695.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'	99.5 / 96.5 / 98.0			
Candidatus Nitrospira nitrificans (GCA_001458775.1)	3'	.	.	G	T	G	C	.	.	.	.	T	.	G	.	A	.	G	.	5'	3'	C	.	T	T	G	T	G	.	C	C	.	.	G	.	.	5'	0.00 / 0.00 / 0.00									
Candidatus Nitrospira nitrosa amoA 1&2 (GCA_001458735.1)	3'	.	.	.	G	C	.	C	.	.	.	C	.	.	C	.	G	.	5'	3'	.	.	G	C	.	.	.	T	.	.	T	.	C	.	.	5'	0.00 / 0.00 / 0.00										
Nitrospira sp. CG24_A (GCA_002869925.1)	3'	C	.	.	.	.	.	C	.	C	T	.	G	.	A	.	G	.	5'	3'	.	G	T	.	.	A	A	A	C	T	.	A	T	C	.	.	5'	0.00 / 0.00 / 0.00									

Figure 3-4. Primer-target mismatch analysis for the newly designed primers and near full-length comammox clade A *amoA* greater than 600 bp. Predicted PCR efficiency is reported separately for the forward and reverse primers in addition to a combined amplification efficiency. Dots indicate a base match and letters indicate a mismatch, with the letter indicating which base is actually present in the sequence.

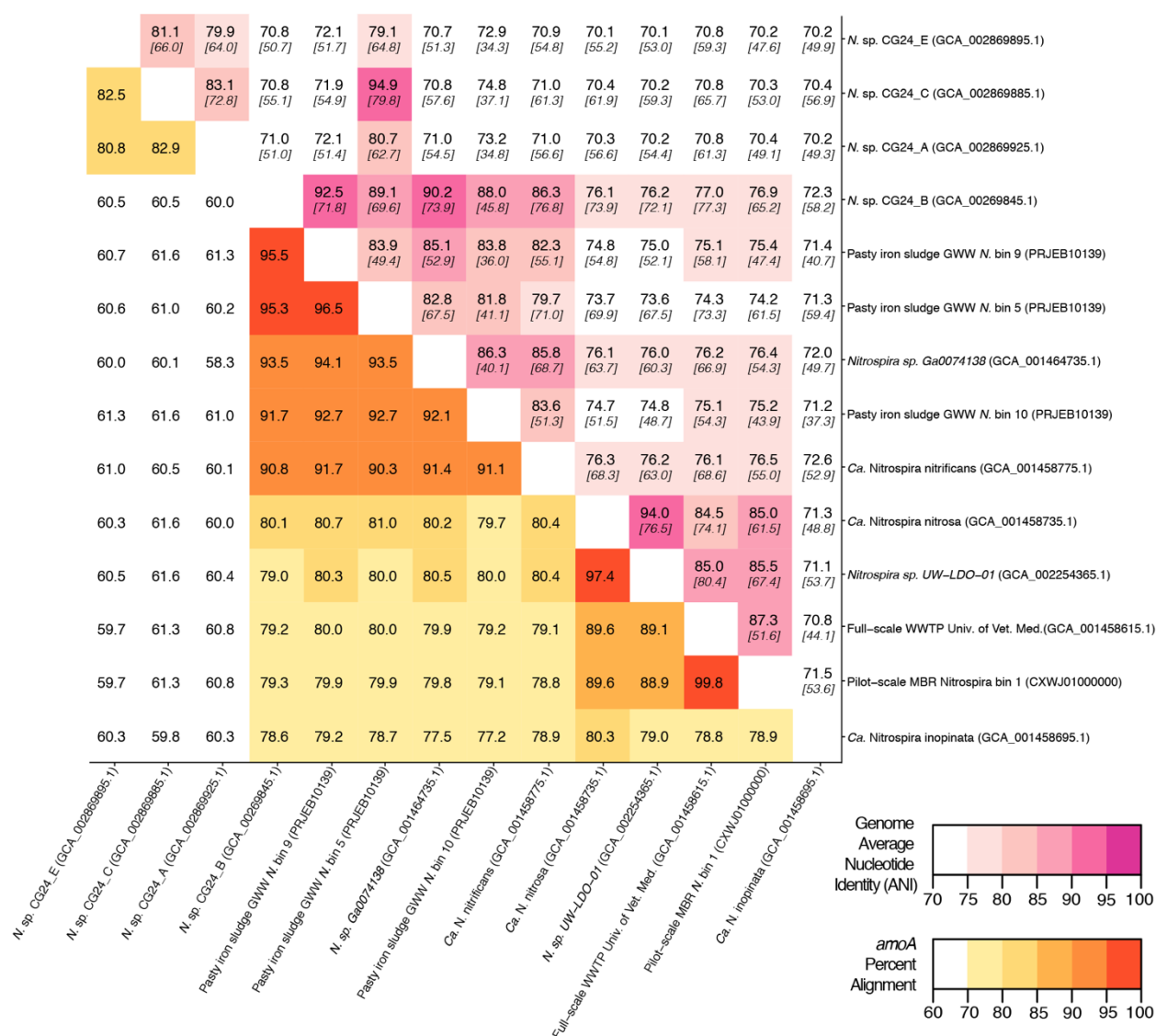


Figure 3-5. Matrix demonstrating the percent alignment of *amoA* genes (white to yellow to red gradient located on the bottom left portion of the matrix) and average nucleotide identity (ANI) of whole genome open reading frame alignments (white to pink gradient located on the top right portion of the matrix). Here, the percent genome ANI is shown in bold on the top of each square and fraction aligned is shown italicized and in brackets beneath the ANI value.

Overall, the *amoA* phylogeny close to *Ca. N. nitrificans* remains unresolved, and therefore, the designed primer set for this species will likely require future refinement.

Sequences from clade B comammox were not included in dataset used for primer design, and therefore, an *in-silico* analysis was also performed with these sequences (Figure 3-4 shows *N. sp. CG24_A amoA* as an example of clade B comammox). In all cases, the sequences had greater than 4 mismatches per primer and are predicted to not amplify with any of the new species-specific primer sets.

The standard curves for all three primer sets were linear (correlation coefficient $R^2 \geq 0.996$) over 7 orders of magnitude (Table S3-1). The amplification efficiency for *Ca. N. nitrosa*, *Ca. N. inopinata*, and *Ca. N. nitrificans amoA* was $96 \pm 3.0\%$, $90 \pm 3.0\%$, and $92 \pm 2.4\%$, respectively (Table 3-3). Melting curve analysis of all standards showed amplification of the target product without primer dimer artifacts, represented by a strong fluorescence signal producing a single melting peak at approximately 88.8 °C for *Ca. N. nitrosa* (Figure S3-10 (A) and S3-11 (A)), 85.6 °C for *Ca. N. inopinata* (Figure S3-12 (A) and S3-13(A)), and 86.7 °C for *Ca. N. nitrificans* (Figure S3-14 (A) and S3-15 (A)).

Environmental detection of candidate comammox species

Using the new species-specific primer sets developed in this study (*Nitrosa amoA*-469f/812r, *Inopinata amoA*-410f/815r, and *Nitrificans amoA*463f/836r), we evaluated the samples originating from the low DO BNR bioreactors (Table 3-1). With the new primers, positive detections obtained with melting curve analysis (Figure S3-10 to S3-15) correlated well with a positive target amplicon in the agarose gel (Figure S3-16 to S3-18). Additionally, cross-hybridization and primer dimers were not observed following PCR with agarose gel

electrophoresis with the standard templates (Figure S3-19). The results from the new species-specific primers are also summarized in decision matrices that were used to consistently evaluate qPCR results between assays (Supplementary Tables S3-6 to S3-8).

Comammox *amoA* belonging to *Ca. N. nitrosa* were detected in all environmental samples. Agarose gel electrophoresis of the qPCR amplified products validated these results, since a single amplicon with the expected length (344-bp) was obtained from all 19 samples (Figure S3-17). *Ca. N. nitrosa* abundance was greater than 10^3 copies *amoA*/ng DNA in nine samples originating from low DO bioreactors L_SBR, P_SBR, P_CF, and TRA (Figure 3-2D). The maximum number of *amoA* copies was obtained from L_SBR-1, with approximately 2.5×10^4 copies *amoA*/ng DNA (Figure 3-2D). After 45 cycles, six samples (L_SBR-4, P_CF-4, NS-1, NS-2, NS-3, and NS-4) were estimated to have fewer copies of *amoA* per ng DNA than the calculated LOD ($M = 11$, 95% CI [9.3, 13] copies/ng DNA) (Figure 3-2D; Table 3-3). However, these positive qPCR detections were confirmed via agarose gel electrophoresis (Figure S3-16A; Figure S3-17), suggesting that the *Ca. N. nitrosa* primers are highly sensitive (Table S3-6). The melting curve analysis at 45 cycles (Figure S3-10) and 30 cycles (Figure S3-11) also confirmed the positive qPCR detections.

Compared to the quantification of total comammox [8], *Ca. N. nitrosa* comprised an average of $340 \pm 70\%$, $96 \pm 29\%$, $67 \pm 13\%$, and 93% of the total comammox population in the true positive detections obtained from L_SBR, P_SBR, P_CF, and TRA, respectively (Figure 3-2A, 2D). Although in some cases the lower quantification with the new species-specific primers could indicate the presence of other non-targeted comammox species, we also see that the total comammox primers contain 4 total mismatches to the primer-target region of *Ca. N. nitrosa*, since mismatches were allowed in the original total comammox primer design

(Table S3-2). In addition, we observed reduced total comammox [8] efficiency when *Ca. N. nitrosa amoA* was present at less than 10^3 copies (Table S3-1). Thus, total comammox [8] primer mismatches to *Ca. N. nitrosa amoA* may contribute to an overall reduced amplification efficiency and underestimation of this species in environmental samples that contain an abundant comammox population, like L_SBR and P_SBR. All in all, the new species-specific primers show that *Ca. N. nitrosa* is an important member of the comammox population in wastewater treatment plants and may contribute to low oxygen nitrification (as seen in L_SBR, P_SBR, P_CF and TRA).

Comammox *amoA* belonging to *Ca. N. inopinata* and *Ca. N. nitrificans* were less frequent and were only detected with less than 10^2 copies *amoA*/ng DNA in all samples (Figure 3-2E, 2F; Figure S3-18). *N. inopinata amoA* was positively detected in one sample from the Madison Metropolitan Sewerage District Nine Springs plant (NS), NS-1 (Table S3-7); although, the concentration of *N. inopinata amoA* was less than the calculated LOD (Figure 3-2E; Table 3-3;). *Ca. N. nitrificans* was only detected above the LOQ in L_SBR-1 and L_SBR-2 with approximately 86 and 21 *amoA*/ng DNA, respectively (Figure 3-2F; Table S3-8). Overall, comammox *amoA* belonging to *Ca. N. inopinata* or *Ca. N. nitrificans* were minor contributors to the total bacteria population after normalization (with relative abundances less than 0.03%) (Figure S3-20).

With the new species-specific primers, we intentionally ran amplifications for 45 cycles, to make sure we could detect any potential non-target amplification. Consequently, the melting curve analysis show some peaks in samples that did not have the target (Figure S3-12 (B-F) and Figure S3-14 (C-F)), but these peaks corresponds to noise after a large number of cycles that would not be seen in quantifications with a more conventional number of

cycles. To verify that the noise in the melting peaks was due to the large number of cycles, we repeated the qPCR experiment with just 30 cycles, which eliminated the peaks that corresponded to noise (Figures S11, S13, S15). Agarose gel electrophoresis of the qPCR amplified products after 30 cycles reaffirm the specificity of the new species-specific primers (Figure S3-16). Moreover, samples that contained noise in the melting peaks when the qPCR was performed with 45 cycles occurred at a late C_q , and consequently, any artifact associated fluorescence was minimal. If these samples had been quantified, the concentrations would be below both the LOD and LOQ, making them easy to eliminate from further analysis without needing to confirm with agarose gel electrophoresis (Table S3-6 to S3-8).

Discussion

Since the late 1800s, nitrification had been described as a division of labor between two phylogenetically distinct and specialized chemolithoautotrophs, the ammonia oxidizers and the nitrite oxidizers [40]. This perception was not challenged for over a century, until Costa et al.,[41] hypothesized that a single nitrifying bacterium combining ammonia oxidation and nitrite oxidation should exist in nature [41]. Nearly one decade later, in late 2015, two separate research groups discovered, cultivated, and characterized the ‘missing’ comammox organism from an aquaculture system and a deep oil exploration well, respectively [1, 2]. To date, there is still uncertainty about the occurrence of the novel *Nitrospira*-like comammox organisms in full scale activated sludge systems. Some studies suggest that comammox are not relevant to conventional wastewater treatment [3, 42]. However, since comammox were discovered in low oxygen environments [1, 2], they may

be able to efficiently utilize oxygen and may be important to low DO BNR, as recently described by Camejo, Santo Domingo [17].

Long-term low DO total nitrogen removal has been successfully demonstrated in laboratory [19] and pilot-scale [19, 20] bioreactors seeded with Nine Springs WWTP activated sludge and operated with DO concentrations less than 0.60 mg O₂/L. Samples analyzed from the Nine Springs WWTP showed presence of *Ca. N. nitrosa*, but below the LOQ. The sludge from this WWTP was the seed for the L_SBR, P_SBR, and P_CF reactors, all of which showed presence of *Ca. N. nitrosa*, suggesting that the low DO conditions favored accumulation of *Ca. N. nitrosa* as a potential participant in ammonia oxidation in low DO BNR. Interestingly, *Ca. N. nitrosa* was also abundant in the full-scale TRA system, which is operated with low DO conditions (Table 3-1), providing independent support for the hypothesis that *Ca. N. nitrosa* is an important contributor to ammonia oxidation in WWTP operated with low DO. *Ca. N. nitrificans* was detected above the LOQ in the L_SBR reactor only, whereas *Ca. N. inopinata* was not detected above the LOD in any of the low DO reactors; therefore, out of the three candidate species, *Ca. N. nitrosa* appears to be the only one that becomes enriched under low DO conditions.

When compared to quantification of the total bacterial population by 16S rRNA-targeted qPCR, the greatest relative abundance of *Ca. N. nitrosa* was 10% and 6% of total bacteria in samples originating from the laboratory-scale and pilot-scale sequencing batch reactors, respectively (Figure S3-20). Although these two reactors were seeded with the same sludge, they were fed from completely different sources (synthetic media for L_SBR and full-scale primary effluent for P_SBR), but operated with low oxygen and a long solids retention time (SRT, both 80 days) [19]. The continuous flow reactors (P_CF and TRA) had lower relative

abundances of *Ca. N. nitrosa*, and although operated with low DO, they had a much shorter SRT (both 10 days; Table 3-1). Thus, in addition to oxygen, SRT may be a factor that contributes to comammox abundance in low DO reactors.

Detection and quantification of microorganisms via real-time PCR relies on designing primers with good specificity to the targeted organisms, good coverage of the targeted group, and good quality of the experimental results. With any newly discovered target group, the quality of primer sets to achieve accurate quantification depends on the quality of the databases used for design and the design considerations. As more sequences of the targeted organisms become available, designs can only improve. The first primer design for comammox quantification described sets for targeting clade A and B *amoA* within the *Nitrospira* genus [18], and subsequently another set of primers was published, aiming for greater coverage to target all comammox *amoA* [8]. Our evaluation of these primer sets in environmental samples from low DO BNR plants showed challenges with primer dimer formation when target sequences were not abundant (clade A and clade B primers [18]), unspecific amplifications in environmental samples (all sets), and underestimation with some strains due to primer-target mismatches (total comammox [8]). These challenges are more important when designing for broad target groups. In general, melting curve analyses of qPCR results are helpful for detection of nonspecific PCR products since different fragments will typically appear as distinct melting peaks [36, 37]. However, the broad detection and quantification of comammox *amoA* has high GC content variability among the comammox organisms described thus far (Table 3-3). Therefore, a wide range of amplicon melting temperatures are expected when using the total comammox [8] or clade-level [18] assays, making melting curve analyses difficult to interpret. We found that

verifying amplified products with the total comammox [8] or clade-level [18] primers cannot be completed with melting curve analysis alone and will require a subsequent agarose gel electrophoresis of our environmental samples, which is not ideal for routine qPCR applications.

Setting the challenge of broad comammox quantification aside, we aimed at designing primer sets specific to the three candidate species described thus far. Evidently, there is a very small number of sequences representative of these candidate species; therefore, the designed primers inherently have 100% coverage. Specificity depends on finding sufficient differences between target and non-targets sequences [24], and the results showed this was possible for the 3 candidate species targeted. As more sequences become available from metagenome assemblies, the specificity of these primers can be re-evaluated. The beginning of such activity was performed in this study (Figure 3-4). Importantly, the qPCR experiments with the newly designed primers were conducted with a longer number of thermal cycles than typical (45 cycles) to increase the chances of detecting potential non-specific amplifications. This extended thermal cycling confirmed that the primers are highly specific and are not producing unwanted non-specific amplifications.

Taken together, the species-specific comammox primers designed in this study enabled an analysis of comammox abundance focused solely on the candidate comammox species currently described in the literature. Although a narrowly focused analysis by design, it eliminated the problems of unspecific amplifications that compromised the use of broad comammox primers and resulted in strong experimental evidence in support of *Ca. N. nitrosa* being a comammox organism in energy-efficient BNR systems operated under low DO conditions. Given earlier studies on low-DO BNR providing some evidence of AOB also

present in these systems [12, 20] and no evidence of AOA [12], we hypothesize that *Ca. N. nitrosa* has an important contribution to ammonia oxidation in low-DO BNR reactors.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

NKB and DRN developed the research plan and project goals. NKB designed the primers and performed laboratory work. Both NKB and DRN drafted the manuscript, tables and figures. DRN contributed sources of project funding. All authors have critically read, corrected, and approved the final version of the manuscript and agree with the opinions expressed here.

Funding

This work was partially supported by funding from the National Science Foundation (CBET-1435661) and the Madison Metropolitan Sewerage District (Madison, WI).

Acknowledgments

This manuscript was previously submitted to the bioRxiv preprint server for biology [43]. We thank Colin Fitzgerald (Jacobs Engineering Group) and Leon Downing (Black & Veatch) for providing collaboration and coordination with the Trinity River Authority (Arlington, TX). We thank Pamela Camejo for providing training for primer design and cloning. We also thank Rahim Ansari, Jonnathan Garcia-Huerta, and Tara Hawes for support with cloning, DNA extraction, and PCR. Finally, we would like to recognize Jackie Bastyr-Cooper for her assistance in the lab with training, protocols and equipment.

References

1. Daims, H., et al., *Complete nitrification by Nitrospira bacteria*. Nature, 2015. **528**(7583): p. 504-9. 10.1038/nature16461
2. van Kessel, M.A., et al., *Complete nitrification by a single microorganism*. Nature, 2015. **528**(7583): p. 555-9. 10.1038/nature16459
3. Chao, Y., et al., *Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach*. Appl Microbiol Biotechnol, 2016. **100**(18): p. 8225-37. 10.1007/s00253-016-7655-9
4. Palomo, A., et al., *Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp.* ISME J, 2016. 10.1038/ismej.2016.63
5. Pinto, A.J., et al., *Metagenomic Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking Water System*. mSphere, 2016. **1**(1): p. e00054-15. 10.1128/mSphere.00054-15
6. Stadler, L.B. and N.G. Love, *Impact of microbial physiology and microbial community structure on pharmaceutical fate driven by dissolved oxygen concentration in nitrifying bioreactors*. Water Res, 2016. **104**: p. 189-199. 10.1016/j.watres.2016.08.001
7. Bartelme, R.P., S.L. McLellan, and R.J. Newton, *Freshwater Recirculating Aquaculture System Operations Drive Biofilter Bacterial Community Shifts around a Stable Nitrifying Consortium of Ammonia-Oxidizing Archaea and Comammox Nitrospira*. Front Microbiol, 2017. **8**: p. 101. 10.3389/fmicb.2017.00101
8. Fowler, S.J., et al., *Comammox Nitrospira are abundant ammonia oxidizers in diverse groundwater-fed rapid sand filter communities*. Environ Microbiol, 2017. **20**(3): p. 1002-1015. 10.1111/1462-2920.14033

9. Hu, H.-W. and J.-Z. He, *Comammox—a newly discovered nitrification process in the terrestrial nitrogen cycle*. *Journal of Soils and Sediments*, 2017. **17**(12): p. 2709-2717. 10.1007/s11368-017-1851-9
10. Park, H.D. and D.R. Noguera, *Evaluating the effect of dissolved oxygen on ammonia-oxidizing bacterial communities in activated sludge*. *Water Res*, 2004. **38**(14-15): p. 3275-86. 10.1016/j.watres.2004.04.047
11. Paredes, D., et al., *New Aspects of Microbial Nitrogen Transformations in the Context of Wastewater Treatment - A Review*. *Engineering in Life Sciences*, 2007. **7**(1): p. 13-25. 10.1002/elsc.200620170
12. Fitzgerald, C.M., et al., *Ammonia-oxidizing microbial communities in reactors with efficient nitrification at low-dissolved oxygen*. *Water Res*, 2015. **70**: p. 38-51. 10.1016/j.watres.2014.11.041
13. Rotthauwe, J.H., K.P. Witzel, and W. Liesack, *The ammonia monooxygenase structural gene amoA as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations*. *Applied and Environmental Microbiology*, 1997. **63**(12): p. 4704-4712. <Go to ISI>://WOS:A1997YK70900013
14. Purkhold, U., et al., *Phylogeny of All Recognized Species of Ammonia Oxidizers Based on Comparative 16S rRNA and amoA Sequence Analysis: Implications for Molecular Diversity Surveys*. *Applied and Environmental Microbiology*, 2000. **66**(12): p. 5368-5382. 10.1128/aem.66.12.5368-5382.2000
15. Koops, H.-P., et al., *The Lithoautotrophic Ammonia-Oxidizing Bacteria*. 2006: p. 778-811. 10.1007/0-387-30745-1_36
16. Pester, M., et al., *amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions*. *Environ Microbiol*, 2012. **14**(2): p. 525-39. 10.1111/j.1462-2920.2011.02666.x

17. Camejo, P.Y., et al., *Genome-Enabled Insights into the Ecophysiology of the Comammox Bacterium "Candidatus Nitrospira nitrosa"*. mSystems, 2017. 2(5). 10.1128/mSystems.00059-17
18. Pjevac, P., et al., *AmoA-Targeted Polymerase Chain Reaction Primers for the Specific Detection and Quantification of Comammox Nitrospira in the Environment*. Front Microbiol, 2017. 8: p. 1508. 10.3389/fmicb.2017.01508
19. Camejo, P.Y., et al., *Candidatus Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors*. Water Res, 2016. 102: p. 125-137. 10.1016/j.watres.2016.06.033
20. Keene, N.A., et al., *Pilot plant demonstration of stable and efficient high rate biological nutrient removal with low dissolved oxygen conditions*. Water Res, 2017. 121(Supplement C): p. 72-85. 10.1016/j.watres.2017.05.029
21. NCBI Resource Coordinators, *Database Resources of the National Center for Biotechnology Information*. Nucleic Acids Res, 2017. 45(D1): p. D12-D17. 10.1093/nar/gkw1071
22. Wright, E.S., *DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment*. BMC Bioinformatics, 2015. 16: p. 322. 10.1186/s12859-015-0749-z
23. Wright, E.S., *Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R*. R Journal, 2016. 8(1): p. 352-359. 10.18129/B9.bioc.DECIPHER
24. Wright, E.S., et al., *Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates*. Environ Microbiol, 2014. 16(5): p. 1354-65. 10.1111/1462-2920.12259
25. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies*. Nucleic Acids Res, 2013. 41(1): p. e1. 10.1093/nar/gks808

26. Thijs, S., et al., *Comparative Evaluation of Four Bacteria-Specific Primer Pairs for 16S rRNA Gene Surveys*. Front Microbiol, 2017. **8**: p. 494. 10.3389/fmicb.2017.00494
27. Palomo, A., et al., *Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox Nitrospira*. ISME J, 2018. 10.1038/s41396-018-0083-3
28. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments*. Clin Chem, 2009. **55**(4): p. 611-22. 10.1373/clinchem.2008.112797
29. Armbruster, D.A. and T. Pry, *Limit of Blank, Limit of Detection and Limit of Quantitation*. The Clinical Biochemist Reviews, 2008. **29**(Suppl 1): p. S49-S52. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556583/>
30. Markowitz, V.M., et al., *IMG: the Integrated Microbial Genomes database and comparative analysis system*. Nucleic Acids Res, 2012. **40**(Database issue): p. D115-22. 10.1093/nar/gkr1044
31. Karst, S.M., R.H. Kirkegaard, and M. Albertsen, *mmgenome: a toolbox for reproducible genome extraction from metagenomes*. bioRxiv, 2016. 10.1101/059121
32. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data*. Bioinformatics, 2012. **28**(12): p. 1647-9. 10.1093/bioinformatics/bts199
33. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification*. BMC Bioinformatics, 2010. **11**: p. 119. 10.1186/1471-2105-11-119
34. Richter, M., et al., *JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison*. Bioinformatics, 2016. **32**(6): p. 929-31. 10.1093/bioinformatics/btv681
35. Camacho, C., et al., *BLAST+: architecture and applications*. BMC Bioinformatics, 2009. **10**: p. 421. 10.1186/1471-2105-10-421

36. Ririe, K.M., R.P. Rasmussen, and C.T. Wittwer, *Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction*. Anal Biochem, 1997. **245**(2): p. 154-160. <https://doi.org/10.1006/abio.1996.9916>
37. Wittwer , C. and N. Kusakawa, *Real-time PCR*, in *Molecular microbiology: diagnostic principles and practice*, D. Persing, et al., Editors. 2004, ASM Press: Washington, DC.
38. Meinhardt, K.A., et al., *Evaluation of revised polymerase chain reaction primers for more inclusive quantification of ammonia-oxidizing archaea and bacteria*. Environ Microbiol Rep, 2015. **7**(2): p. 354-63. 10.1111/1758-2229.12259
39. Richter, M. and R. Rossello-Mora, *Shifting the genomic gold standard for the prokaryotic species definition*. Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19126-31. 10.1073/pnas.0906412106
40. Brock, T.D., *Milestones in Microbiology 1546 to 1940*. American Society Mic Series. 1999: ASM Press. <https://books.google.com/books?id=q5JHcs8w21gC>
41. Costa, E., J. Perez, and J.U. Kreft, *Why is metabolic labour divided in nitrification?* Trends Microbiol, 2006. **14**(5): p. 213-9. 10.1016/j.tim.2006.03.006
42. Gonzalez-Martinez, A., et al., *Detection of comammox bacteria in full-scale wastewater treatment bioreactors using tag-454-pyrosequencing*. Environ Sci Pollut Res Int, 2016. **23**(24): p. 25501-25511. 10.1007/s11356-016-7914-4
43. Keene-Beach, N. and D.R. Noguera, *Design and assessment of species-level qPCR primers targeting comammox*. bioRxiv, 2018. 10.1101/348664

Supplementary Tables

Table S3-1. qPCR standard curves: slope, y-intercept, correlation coefficient, and PCR efficiency calculated from the approximate linear range for each assay included in this study.

	Primer set	Log ₁₀ linear range (log(SQ))*	Std curve slope (C _q /log(SQ))*	Std curve Y-intercept (C _q)*	Std curve correlation coefficient (R ²)	PCR efficiency (%)
Comammox <i>Nitrospira</i> clade A	Standard template: <i>Ca. N. nitrosa amoA</i>					
	Nitrosa amoA-469F/812R ^a	1 - 7	-3.44 ± 0.07	37.6 ± 0.56	0.998 ± 0.001	96 ± 3.0
	Ntsp-amoA 162F/359R ^b	3 - 7	-3.55 ± 0.07	38.5 ± 0.43	0.999 ± 0.001	91 ± 2.6
	comaA-244f/659r ^c	2 - 7	-3.49 ± 0.07	37.0 ± 0.49	0.998 ± 0.001	95 ± 2.7
	Standard template: <i>Ca. N. inopinata amoA</i>					
	Inopinata amoA-410F/815R ^a	1 - 7	-3.63 ± 0.13	38.0 ± 0.65	0.997 ± 0.002	90 ± 3.0
	Ntsp-amoA 162F/359R ^b	2 - 7	-3.84 ± 0.12	39.5 ± 0.64	0.998 ± 0.002	82 ± 3.4
	comaA-244f/659r ^c	3 - 7	-3.70 ± 0.08	37.3 ± 0.45	0.999 ± 0.001	86 ± 2.5
	Standard template: <i>Ca. N. nitrificans amoA</i>					
	Nitrificans amoA-463F/836R ^a	1 - 7	-3.52 ± 0.07	37.8 ± 0.11	0.996 ± 0.004	92 ± 2.4
	Ntsp-amoA 162F/359R ^b	2 - 7	-3.43 ± 0.05	36.9 ± 0.27	0.999 ± 0.001	96 ± 1.8
	comaA-244f/659r ^c	2 - 7	-3.51 ± 0.08	36.2 ± 0.56	0.996 ± 0.006	93 ± 2.9
Comammox <i>Nitrospira</i> clade B	Standard template: <i>N. CG24_E amoA</i>					
	comaB-244f/659r ^c	1 - 7	-3.76 ± 0.25	38.5 ± 0.23	0.992 ± 0.005	88 ± 0.6
Total 16S rRNA	Standard template: <i>Actinobacteria</i> (ac1) 16S rRNA					
	16S-341f/785r ^d	2 - 7	-3.88 ± 0.03	39.4 ± 0.15	0.997 ± 0.001	81 ± 0.8

*Where C_q = quantification cycle; SQ = starting quantity in copies per reaction; Std = standard

a. This study

b. Fowler et al., 2017

c. Pjevac et al., 2017

d. Klindworth et al., 2013; Thijs et al., 2017

Table S3-2. Total comammox primer mismatch table with comammox standard template *amoA* used in this study

		Forward Target Sequence																				
Ntsp-amoA 162F	5'	G	G	A	T	T	T	C	T	G	G	N	T	S	G	A	T	T	G	G	A	5'
Ca. N. nitrosa amoA1 (GCA_001458735.1)	3'	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	5'
Ca. N. nitrosa amoA2 (GCA_001458735.1)	3'	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	.	.	.	5'
Ca. N. inopinata amoA (GCA_001458695.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'
Ca. N. nitrificans amoA (GCA_001458775.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5'
		Reverse Target Sequence																				
Ntsp-amoA 359R	5'	W	A	G	T	T	N	G	A	C	C	A	C	C	A	S	T	A	C	C	A	3'
Ca. N. nitrosa amoA1 (GCA_001458735.1)	3'	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	5'
Ca. N. nitrosa amoA2 (GCA_001458735.1)	3'	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	5'
Ca. N. inopinata amoA (GCA_001458695.1)	3'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	5'
Ca. N. nitrificans amoA (GCA_001458775.1)	3'	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	5'

Table S3-3. Total comamox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

Total comamox assay ^a				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Bioreactor Sample ID	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	Yes	No	20	Yes	Yes	Yes
L_SBR-2	3/18/13	Yes	No	21	Yes	Yes	Yes
L_SBR-3	4/17/13	No	Yes	24	Yes	Yes	Yes
L_SBR-4	10/3/14	No	No	22	Yes	Yes	No
P_SBR-1	9/10/15	Yes	No	19	Yes	Yes	Yes
P_SBR-2	11/10/15	Yes	No	18	Yes	Yes	Yes
P_SBR-3	12/15/15	Yes	No	18	Yes	Yes	Yes
P_SBR-4	4/25/16	Yes	Yes	18	Yes	Yes	No
P_CF-1	9/9/15	Yes	No	24	Yes	Yes	Yes
P_CF-2	11/18/15	Yes	No	21	Yes	Yes	Yes
P_CF-3	12/30/15	Yes	Yes	22	Yes	Yes	Yes
P_CF-4	4/26/16	No	Yes	20	Yes	Yes	Yes
TRA-1	8/9/16	Yes	No	21	Yes	Yes	Yes
TRA-2	7/28/17	Yes	No	22	Yes	Yes	Yes
TRA-3	7/28/17	Yes	No	21	Yes	Yes	Yes
NS-1	9/9/15	Yes	Yes	22	Yes	Yes	No
NS-2	11/18/15	Yes	Yes	22	Yes	Yes	No
NS-3	12/30/15	Yes	Yes	22	Yes	Yes	Yes
NS-4	4/26/16	Yes	Yes	21	Yes	Yes	Yes

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^a Fowler et al. 2017

Table S3-4. Clade A comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

Clade A assay ^b				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Reactor	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	Yes	No	18	Yes	Yes	Yes
L_SBR-2	3/18/13	Yes	No	19	Yes	Yes	Yes
L_SBR-3	4/17/13	Yes	No	22	Yes	Yes	Yes
L_SBR-4	10/3/14	No	Yes	22	Yes	Yes	No
P_SBR-1	9/10/15	Yes	Yes	18	Yes	Yes	Yes
P_SBR-2	11/10/15	Yes	Yes	16	Yes	Yes	Yes
P_SBR-3	12/15/15	Yes	No	17	Yes	Yes	No
P_SBR-4	4/25/16	No	Yes	19	Yes	Yes	No
P_CF-1	9/9/15	No	No	22	Yes	Yes	Yes
P_CF-2	11/18/15	No	Yes	19	Yes	Yes	No
P_CF-3	12/30/15	No	Yes	20	Yes	Yes	Yes
P_CF-4	4/26/16	Yes	Yes	19	Yes	Yes	No
TRA-1	8/9/16	No	Yes	21	Yes	Yes	Yes
TRA-2	7/28/17	No	Yes	22	Yes	Yes	No
TRA-3	7/28/17	Yes	Yes	20	Yes	Yes	Yes
NS-1	9/9/15	Yes	Yes	22	Yes	Yes	Yes
NS-2	11/18/15	Yes	Yes	21	Yes	Yes	Yes
NS-3	12/30/15	Yes	Yes	20	Yes	Yes	Yes
NS-4	4/26/16	Yes	Yes	20	Yes	Yes	Yes

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^b Pjevac et al. 2017

Table S3-5. Clade B comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

Clade B assay ^b				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Reactor	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	No	Yes	22	Yes	Yes	No
L_SBR-2	3/18/13	No	Yes	23	Yes	Yes	No
L_SBR-3	4/17/13	No	No	23	Yes	Yes	No
L_SBR-4	10/3/14	No	Yes	23	Yes	Yes	No
P_SBR-1	9/10/15	No	Yes	19	Yes	Yes	Yes
P_SBR-2	11/10/15	No	Yes	19	Yes	Yes	Yes
P_SBR-3	12/15/15	No	Yes	19	Yes	Yes	No
P_SBR-4	4/25/16	No	Yes	19	Yes	Yes	No
P_CF-1	9/9/15	No	Yes	22	Yes	Yes	Yes
P_CF-2	11/18/15	No	Yes	20	Yes	Yes	Yes
P_CF-3	12/30/15	No	Yes	21	Yes	Yes	Yes
P_CF-4	4/26/16	Questionable	Yes	19	Yes	Yes	No
TRA-1	8/9/16	No	Yes	21	Yes	Yes	No
TRA-2	7/28/17	No	Yes	21	Yes	Yes	No
TRA-3	7/28/17	No	Yes	21	Yes	Yes	No
NS-1	9/9/15	No	Yes	23	Yes	Yes	No
NS-2	11/18/15	No	Yes	22	Yes	Yes	No
NS-3	12/30/15	No	No	22	Yes	Yes	No
NS-4	4/26/16	No	Yes	21	Yes	Yes	No

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^b Pjevac et al. 2017

Table S3-6. Ca. *N. nitrosa* comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

N. nitrosa assay ^c				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Reactor	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	Yes	No	17	Yes	Yes	Yes
L_SBR-2	3/18/13	Yes	No	19	Yes	Yes	Yes
L_SBR-3	4/17/13	Yes	No	27	Yes	Yes	Yes
L_SBR-4	10/3/14	Yes, after 30 cycles	No	34	No	No	Yes
P_SBR-1	9/10/15	Yes	No	18	Yes	Yes	Yes
P_SBR-2	11/10/15	Yes	No	17	Yes	Yes	Yes
P_SBR-3	12/15/15	Yes	No	18	Yes	Yes	Yes
P_SBR-4	4/25/16	Yes	No	21	Yes	Yes	Yes
P_CF-1	9/9/15	Yes	No	24	Yes	Yes	Yes
P_CF-2	11/18/15	Yes	No	21	Yes	Yes	Yes
P_CF-3	12/30/15	Yes	No	22	Yes	Yes	Yes
P_CF-4	4/26/16	Yes	No	29	No	No	Yes
TRA-1	8/9/16	Yes	No	21	Yes	Yes	Yes
TRA-2	7/28/17	Yes	No	25	Yes	Yes	Yes
TRA-3	7/28/17	Yes	No	20	Yes	Yes	Yes
NS-1	9/9/15	Yes	No	29	No	No	Yes
NS-2	11/18/15	Yes	No	29	No	No	Yes
NS-3	12/30/15	Yes	No	29	No	No	Yes
NS-4	4/26/16	Yes, after 30 cycles	No	33	No	No	Yes

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^c This Study

Table S3-7. *Ca. N. inopinata* comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

N. inopinata assay ^c				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Reactor	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	No	No	39	No	No	No
L_SBR-2	3/18/13	No	No	40	No	No	No
L_SBR-3	4/17/13	No	No	40	No	No	No
L_SBR-4	10/3/14	No	No	34	No	No	No
P_SBR-1	9/10/15	No	No	38	No	No	No
P_SBR-2	11/10/15	No	No	40	No	No	No
P_SBR-3	12/15/15	No	No	40	No	No	No
P_SBR-4	4/25/16	No	No	38	No	No	No
P_CF-1	9/9/15	No	No	39	No	No	No
P_CF-2	11/18/15	No	No	39	No	No	No
P_CF-3	12/30/15	No	No	39	No	No	No
P_CF-4	4/26/16	No	No	37	No	No	No
TRA-1	8/9/16	No	No	38	No	No	No
TRA-2	7/28/17	No	No	39	No	No	No
TRA-3	7/28/17	No	No	39	No	No	No
NS-1	9/9/15	Yes, after 30 cycles	No	33	No	No	Yes
NS-2	11/18/15	No	No	40	No	No	No
NS-3	12/30/15	No	No	39	No	No	No
NS-4	4/26/16	No	No	39	No	No	No

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^c This Study

Table S3-8. *Ca. N* nitrificans comammox qPCR assay decision matrix summarizing criteria used to determine which samples were a non-detect, true positive, overestimate, or false positive

N. nitrificans assay ^c				
	Non-detect	True Positive	Overestimate	False Positive
Target band in gel?	No	Yes	Yes	No
Off target bands?	No	No	Yes	-
SQ > LOD *	No	Yes	Yes	Yes
SQ > LOQ *	No	Yes or No	Yes	Yes or No

Reactor	Date	Qualitative		Quantitative			
		Target band in gel?	Off target bands?	Average C _q *	SQ > LOD*	SQ > LOQ*	Average T _m * within limits?
L_SBR-1	3/7/13	Yes	No	25	Yes	Yes	Yes
L_SBR-2	3/18/13	Yes	No	27	Yes	Yes	Yes
L_SBR-3	4/17/13	Yes	No	29	Yes	No	Yes
L_SBR-4	10/3/14	Yes, after 30 cycles	No	35	No	No	Yes
P_SBR-1	9/10/15	Yes, after 30 cycles	No	30	Yes	No	No
P_SBR-2	11/10/15	No	No	31	No	No	No
P_SBR-3	12/15/15	No	No	31	No	No	No
P_SBR-4	4/25/16	No	No	31	No	No	No
P_CF-1	9/9/15	No	No	33	No	No	No
P_CF-2	11/18/15	No	No	32	No	No	No
P_CF-3	12/30/15	No	No	31	No	No	No
P_CF-4	4/26/16	No	No	32	No	No	No
TRA-1	8/9/16	No	No	32	No	No	No
TRA-2	7/28/17	No	No	31	No	No	No
TRA-3	7/28/17	No	No	31	No	No	No
NS-1	9/9/15	No	No	33	No	No	No
NS-2	11/18/15	No	No	33	No	No	No
NS-3	12/30/15	No	No	31	No	No	No
NS-4	4/26/16	No	No	32	No	No	No

* C_q = quantification cycle, SQ = starting quantity, LOD = limit of detection, LOQ = limit of quantification, T_m = melting temperature

^c This Study

Supplementary Figures

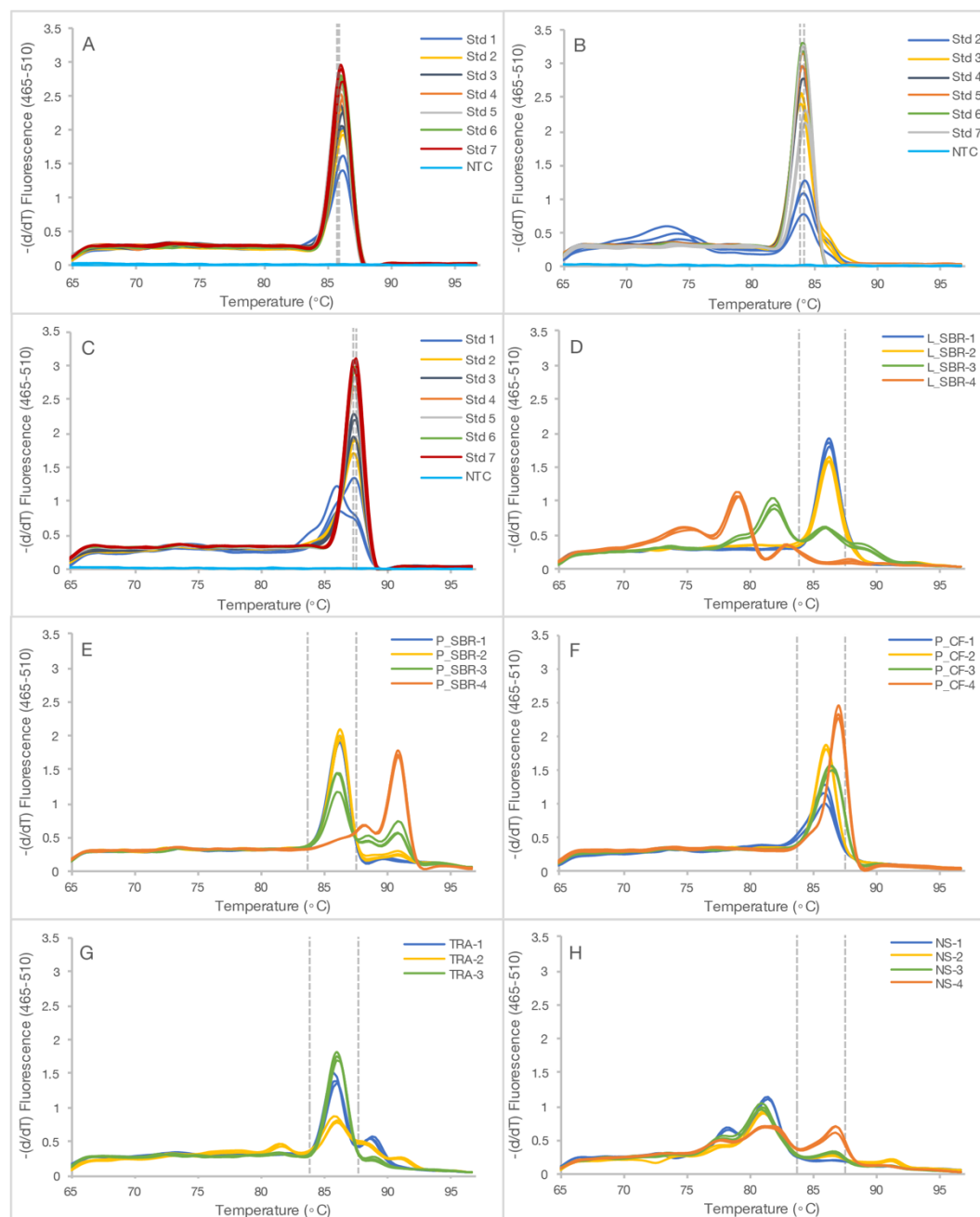


Figure S3-1. Total comammox assay (Ntsp-amoA 162F/359R) melting peaks for A) *Ca. N. nitrosa* amoA standard curve, B) *Ca. N. inopinata* amoA standard curve, and C) *Ca. N. nitrificans* amoA standard curve, D) L-SBR samples, E) P-SBR samples, F) P-CF samples, G) TRA samples, and H) NS samples. The dashed lines in panels A-C represent the standard deviation of melting temperatures for that specific standard. The dashed lines in panels D-H represent the minimum and maximum range of melting temperatures derived from all the standard curves. Triplicate data series are shown for each sample.

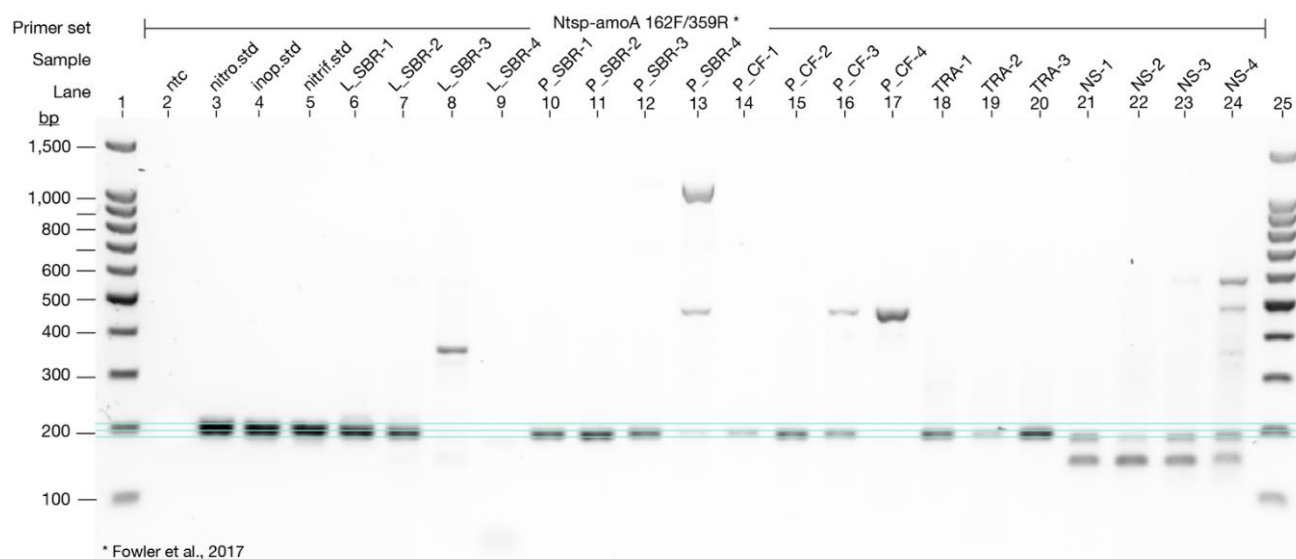


Figure S3-2. Agarose gel electrophoresis was performed using the total comammox (Ntsp-amoA 162F/359R) assay and included one replicate from each of the following samples: a non-template control (ntc), *Ca. N. nitrosa* amoA standard (nitro.std), *Ca. N. inopinata* amoA standard (inop.std), *Ca. N. nitrificans* amoA standard (nitrif.std), and one replicate from each of the 19 bioreactor samples. The expected amplicon size is 198 bp for this assay.

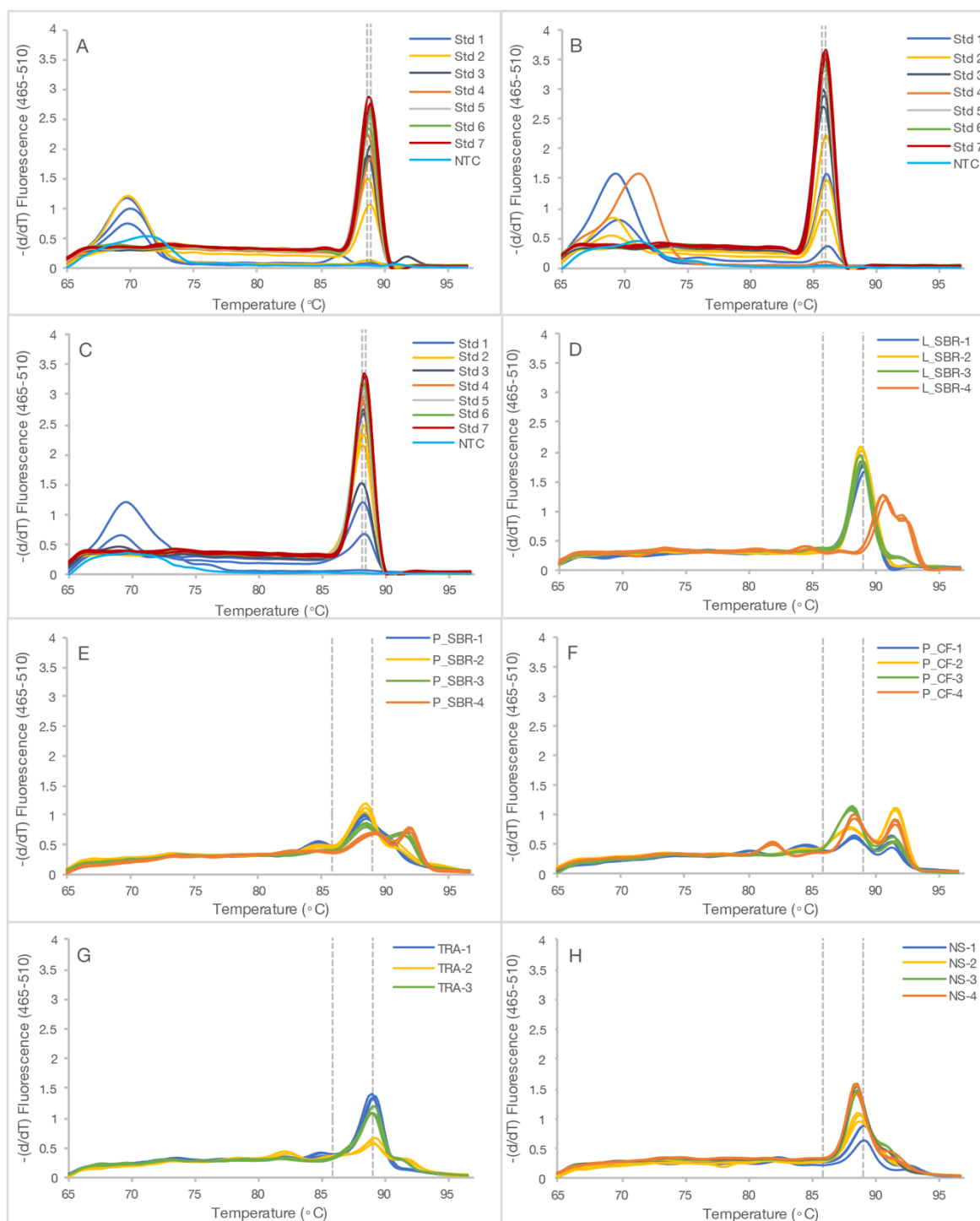


Figure S3-3. Clade A comammox assay (coma-244f/659r) melting peaks for A) *Ca. N. nitrosa amoA* standard curve, B) *Ca. N. inopinata amoA* standard curve, C) *Ca. N. nitrificans amoA* standard curve, D) L-SBR samples, E) P-SBR samples, F) P-CF samples, G) TRA samples, and H) NS samples. The dashed lines in panels A-C represent the standard deviation of melting temperatures for that specific standard. The dashed lines in panels D-H represent the minimum and maximum range of melting temperatures derived from all the standard curves. Triplicate data series are shown for each sample.

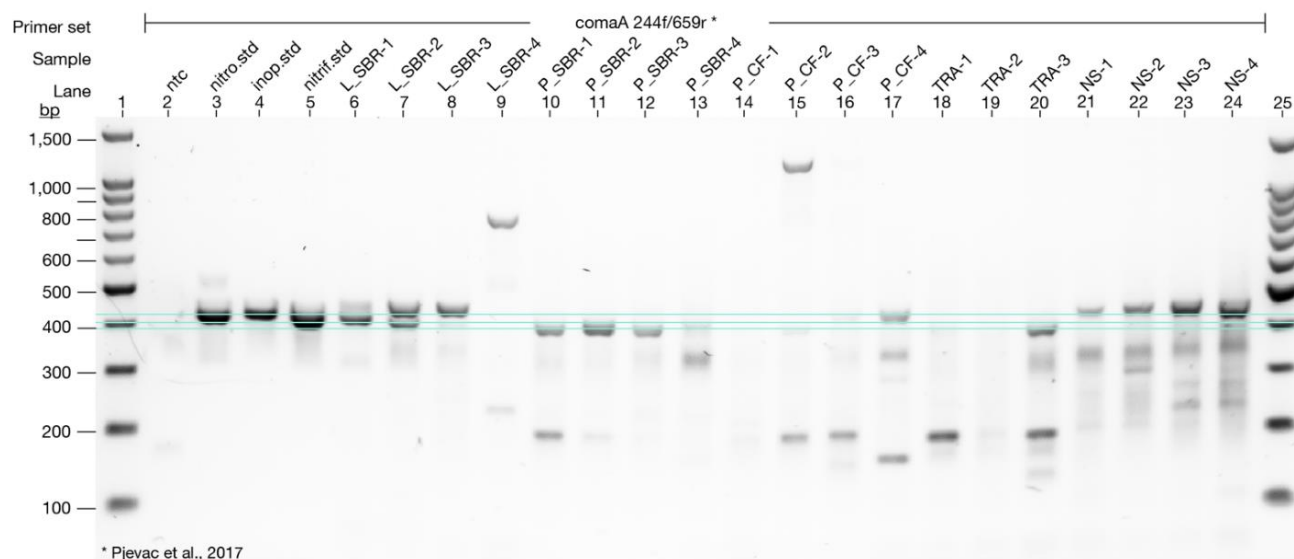


Figure S3-4. Agarose gel electrophoresis was performed using the clade A comammox (comA-244f/659r) assay and included one replicate from each of the following samples: a non-template control (ntc), *Ca. N. nitrosa* amoA standard (nitro.std), *Ca. N. inopinata* amoA standard (inop.std), *Ca. N. nitrificans* amoA standard (nitrif.std), and one replicate from each of the 19 bioreactor samples. The expected amplicon size for this assay is 415 bp.

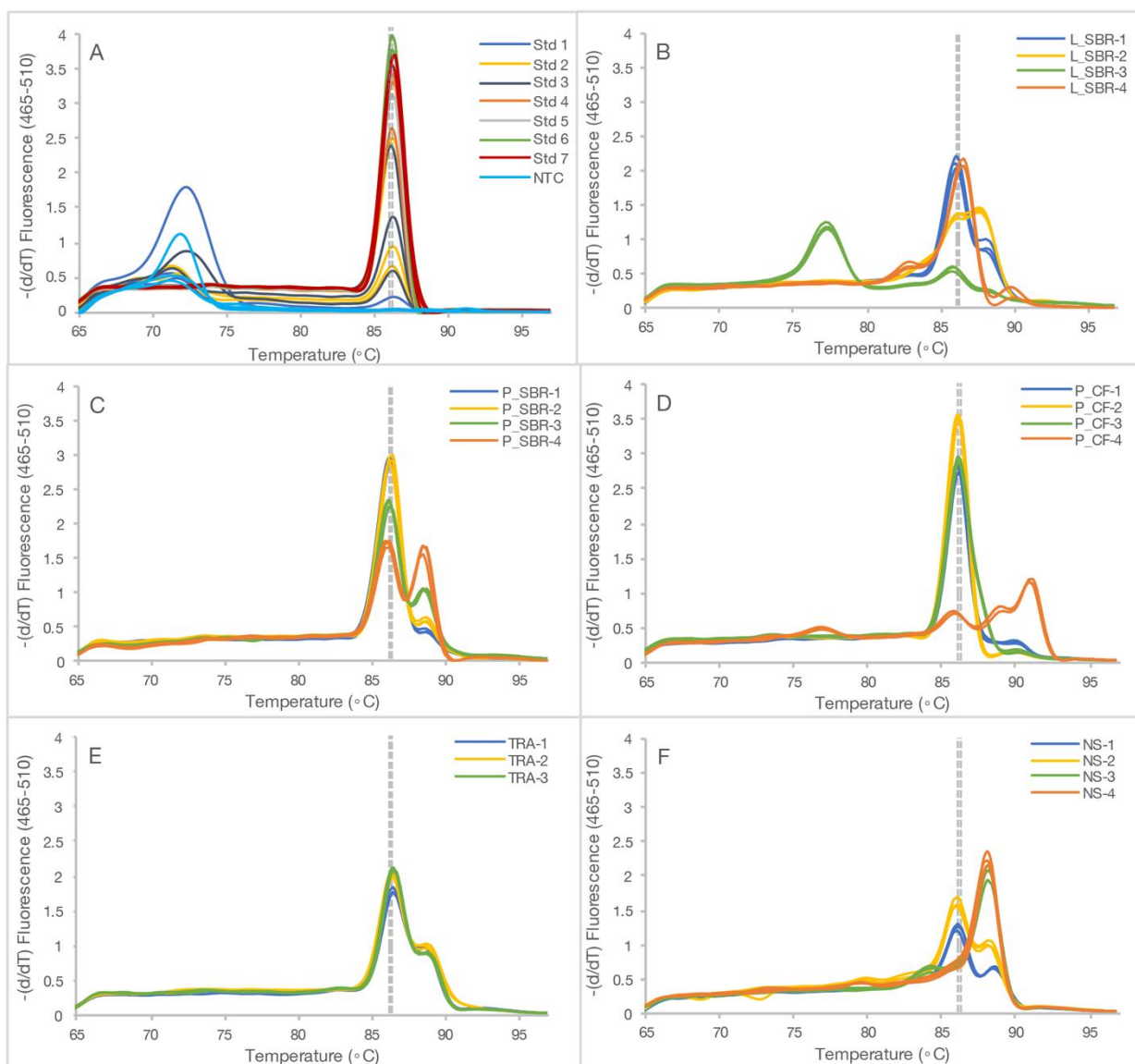


Figure S3-5. Clade B comammox assay (coma-244f/659r) melting peaks for A) *Nitrospira* sp CG24_E *amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the standard deviation of melting temperatures for the standard. Triplicate data series are shown for each sample.

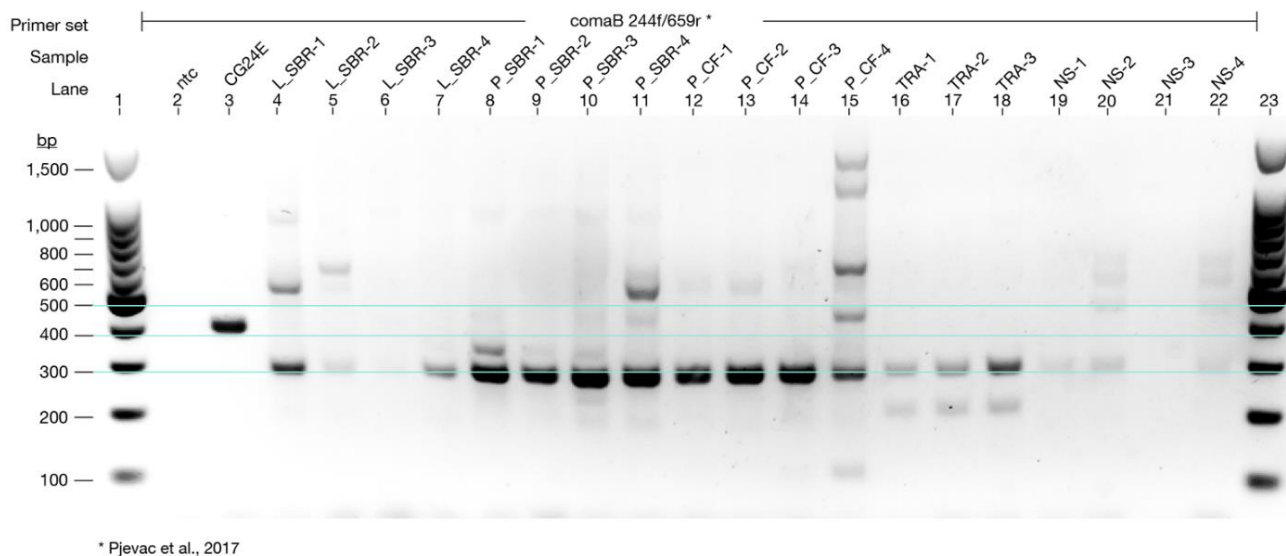


Figure S3-6. Agarose gel electrophoresis was performed using the clade B comammox (comaB-244f/659r) assay and included one replicate from each of the following samples: a non-template control (ntc), *Nitrospira* sp. CG24_E *amoA* standard (CG24E), and one replicate from each of the 19 bioreactor samples. The expected amplicon size for this assay is 415 bp

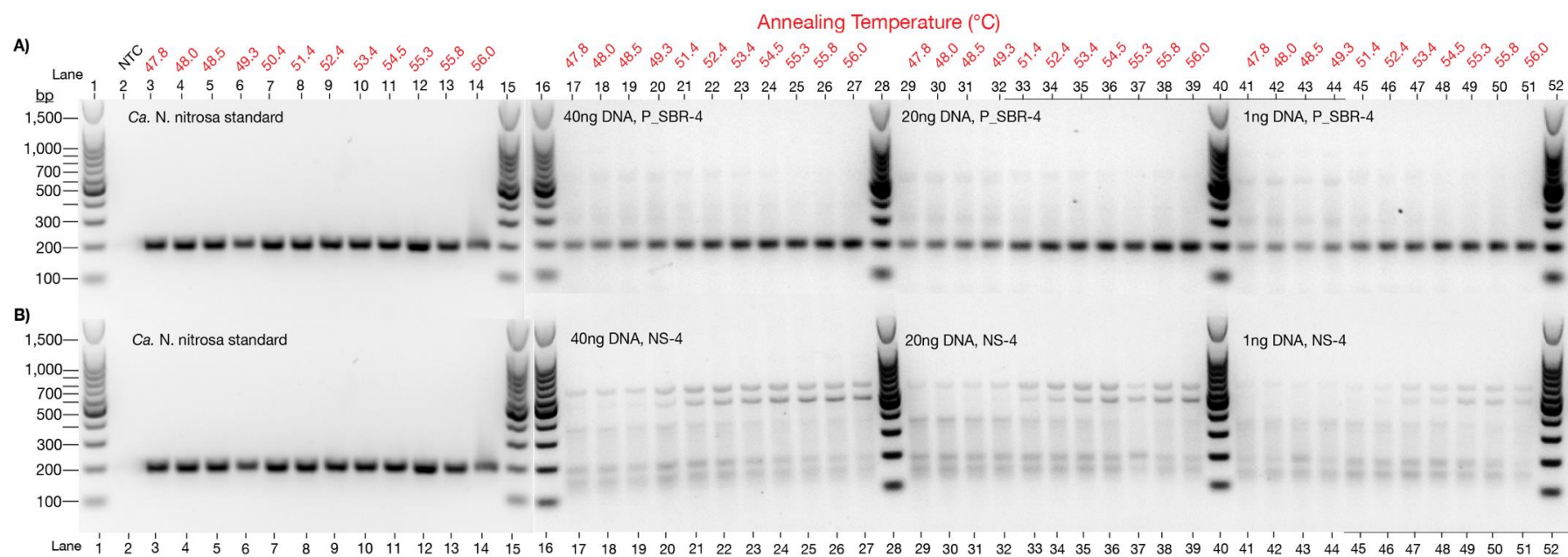


Figure S3-7. Total comammox (Ntsp-amoA 162F/359R) gradient PCR assay results with *Ca. N. nitrosa* standard and two bioreactor samples (A) P_SBR-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 47.8 °C and 56.0 °C, where 48 °C is the original design annealing temperature. The expected amplicon size is 198 bp for this assay. Off-target products are observed at approximately 300 bp and 600 bp in the P_SBR-4 sample and at approximately 150 bp, 350 bp, 600 bp and 700 bp in the NS-4 sample, and appear to be independent of the total DNA concentrations tested in the reaction.

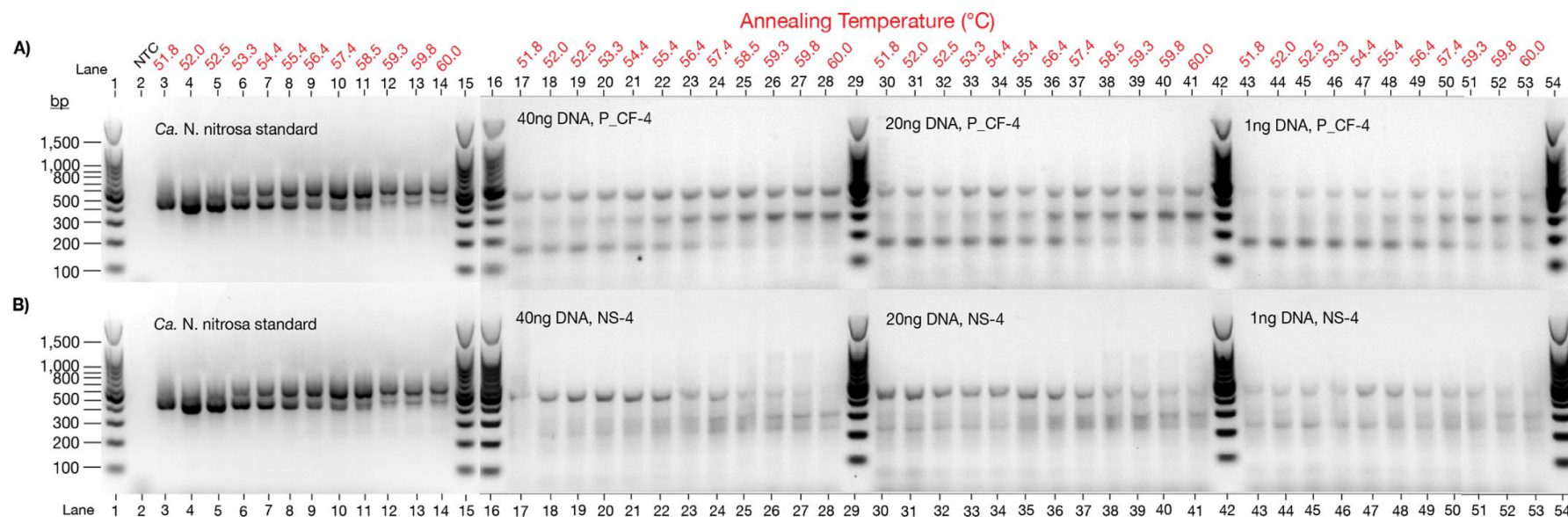


Figure S3-8. Clade A comammox (comaA-244f/659r) gradient PCR assay results with *Ca. N. nitrosa* standard and two bioreactor samples (A) P_CF-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 51.8 °C and 60.0 °C, where 52 °C is the original design annealing temperature. The expected amplicon size for this assay is 415 bp. Off-target products are observed at approximately 150 bp and 300 bp in the P_CF-4 sample and at approximately 250 bp and 300 bp in the NS-4 sample, and the amplification of these off-target products appear to be independent of the total DNA concentrations tested in the reaction. That is, with the exception of the P_CF-4 sample at the lowest total DNA concentration tested, where the 150 bp off-target product appears brighter than the target at the design annealing temperature, indicating the off-target product may be selectively amplified over the target in these conditions.

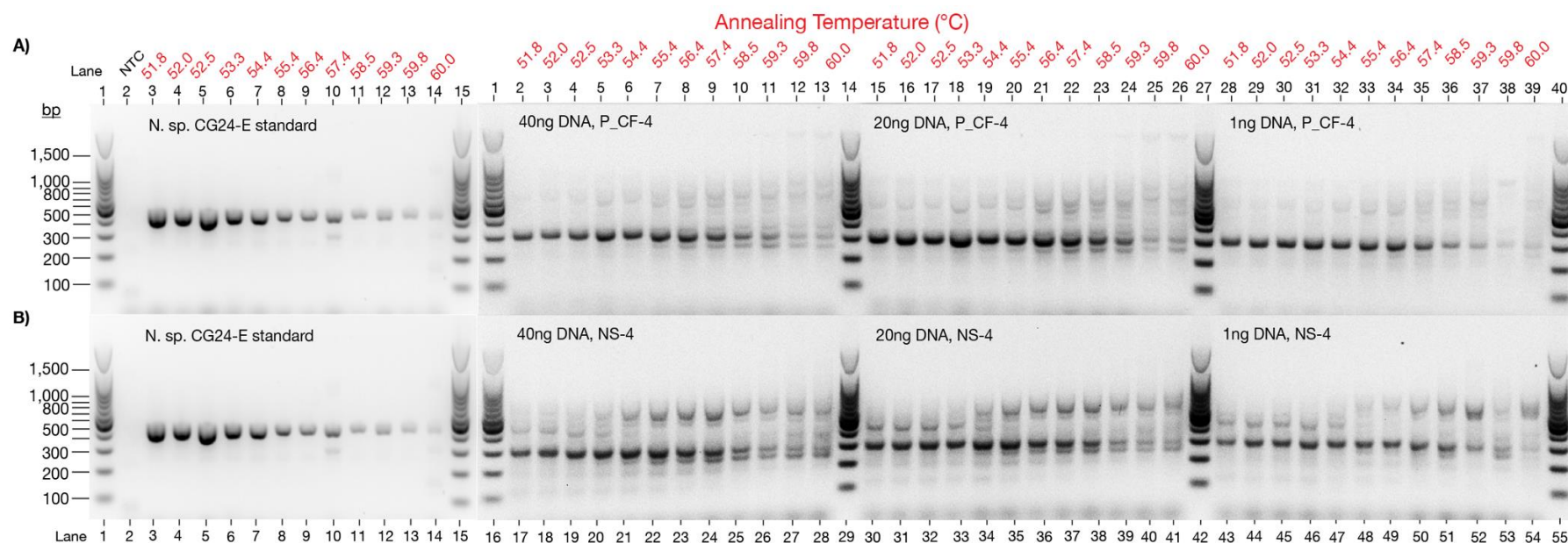


Figure S3-9. Clade B comammox (comaB-244f/659r) gradient PCR assay results with *Nitrospira* sp. CG-24E standard and two bioreactor samples (A) P_CF-4 and (B) NS-4, each diluted to 40ng DNA/reaction, 20ng DNA/reaction, and 1ng DNA/reaction. The gradient PCR was performed with annealing temperatures ranging between 51.8 °C and 60.0 °C, where 52 °C is the original design annealing temperature. The expected amplicon size for this assay is 415 bp. Multiple off-target products, especially one at 300 bp, appear at nearly every DNA concentration and temperature tested.

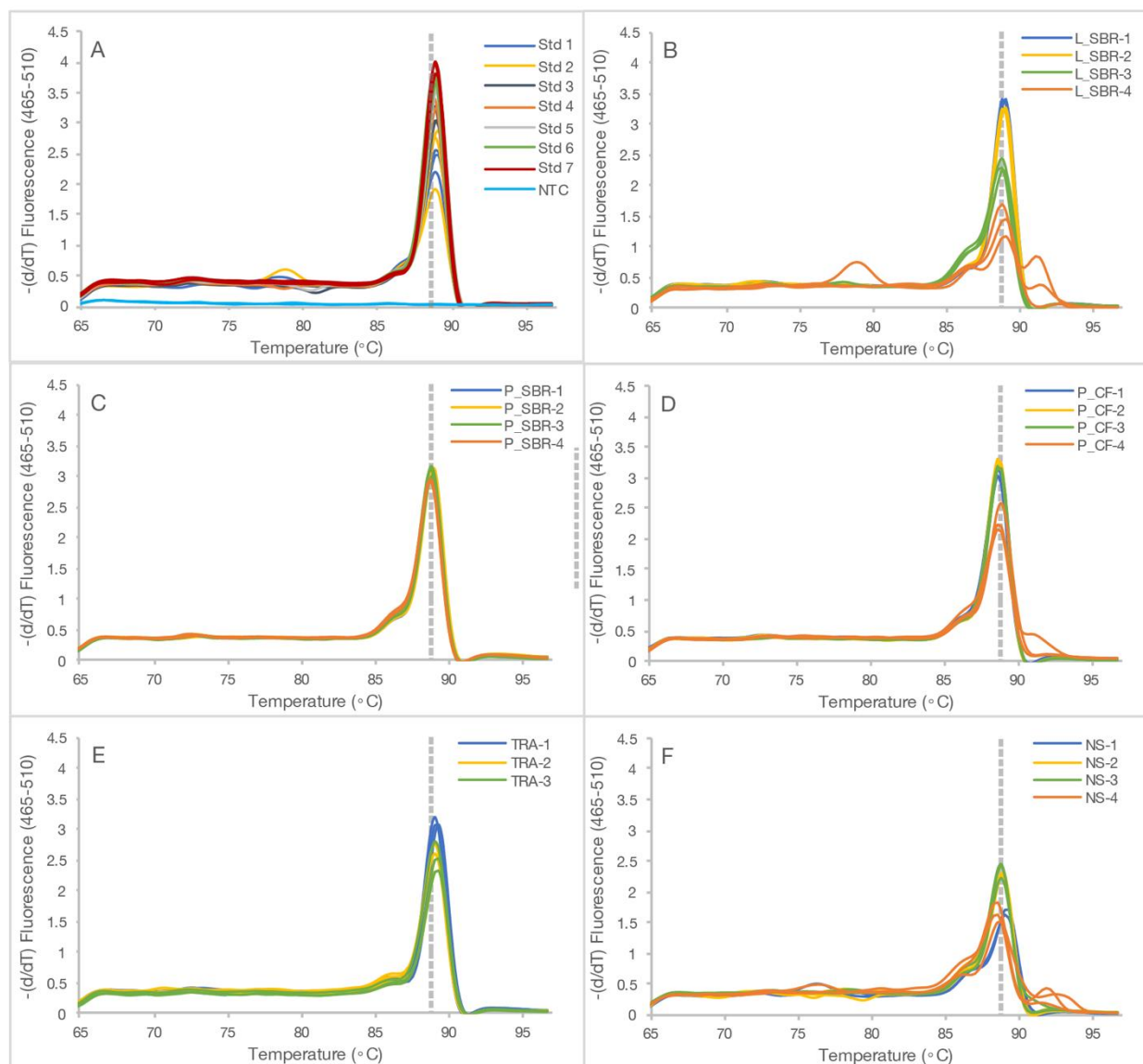


Figure S3-10. Comammox *Ca. N. nitrosa* assay (Nitrosa amoA-469F/812R) melting peaks after 45 cycles for A) *N. nitrosa* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample.

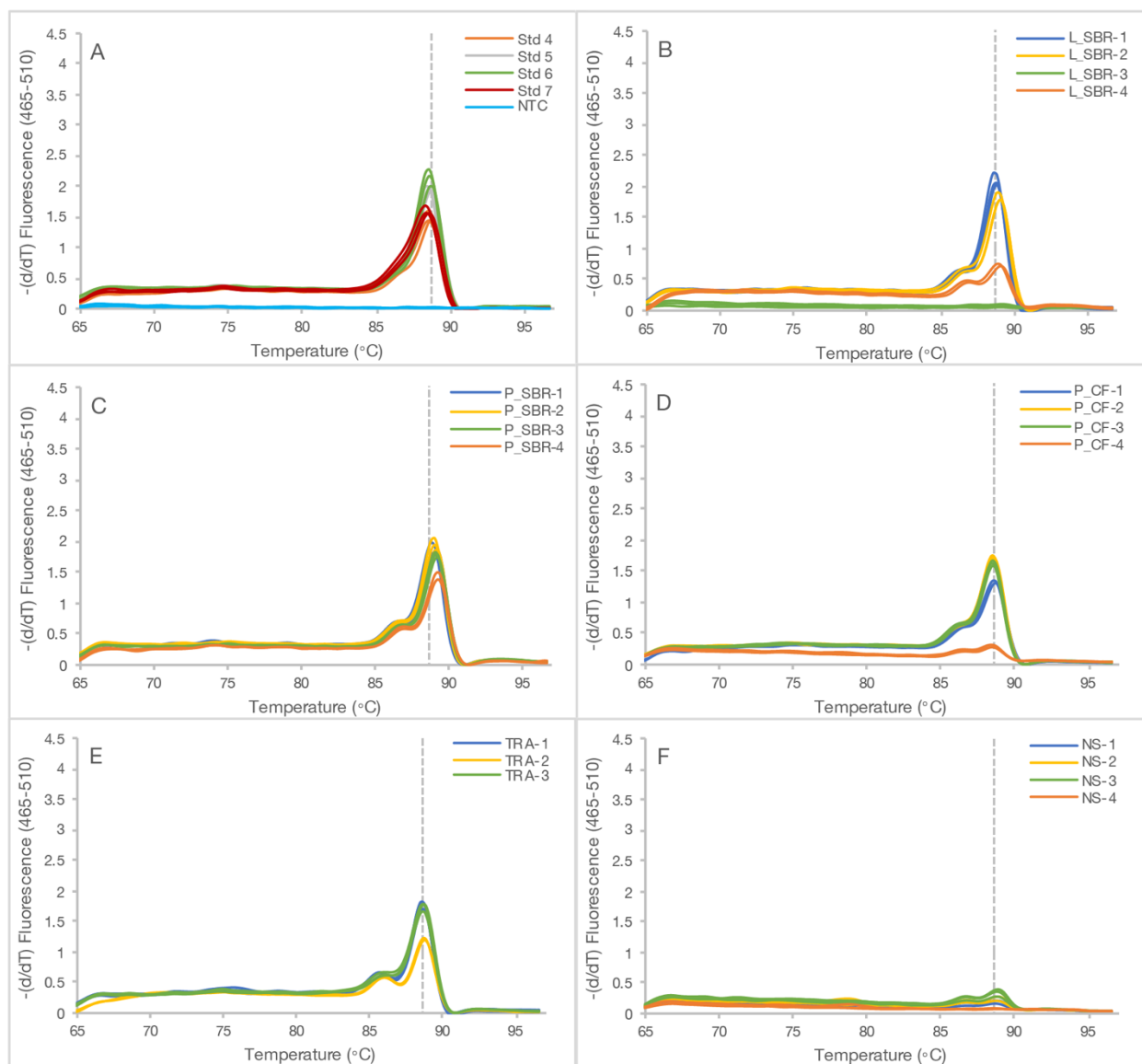


Figure S3-11. Comammox *Ca. N. nitrosa* assay (Nitrosa amoA-469F/812R) melting peaks after 30 cycles for A) *N. nitrosa* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample.

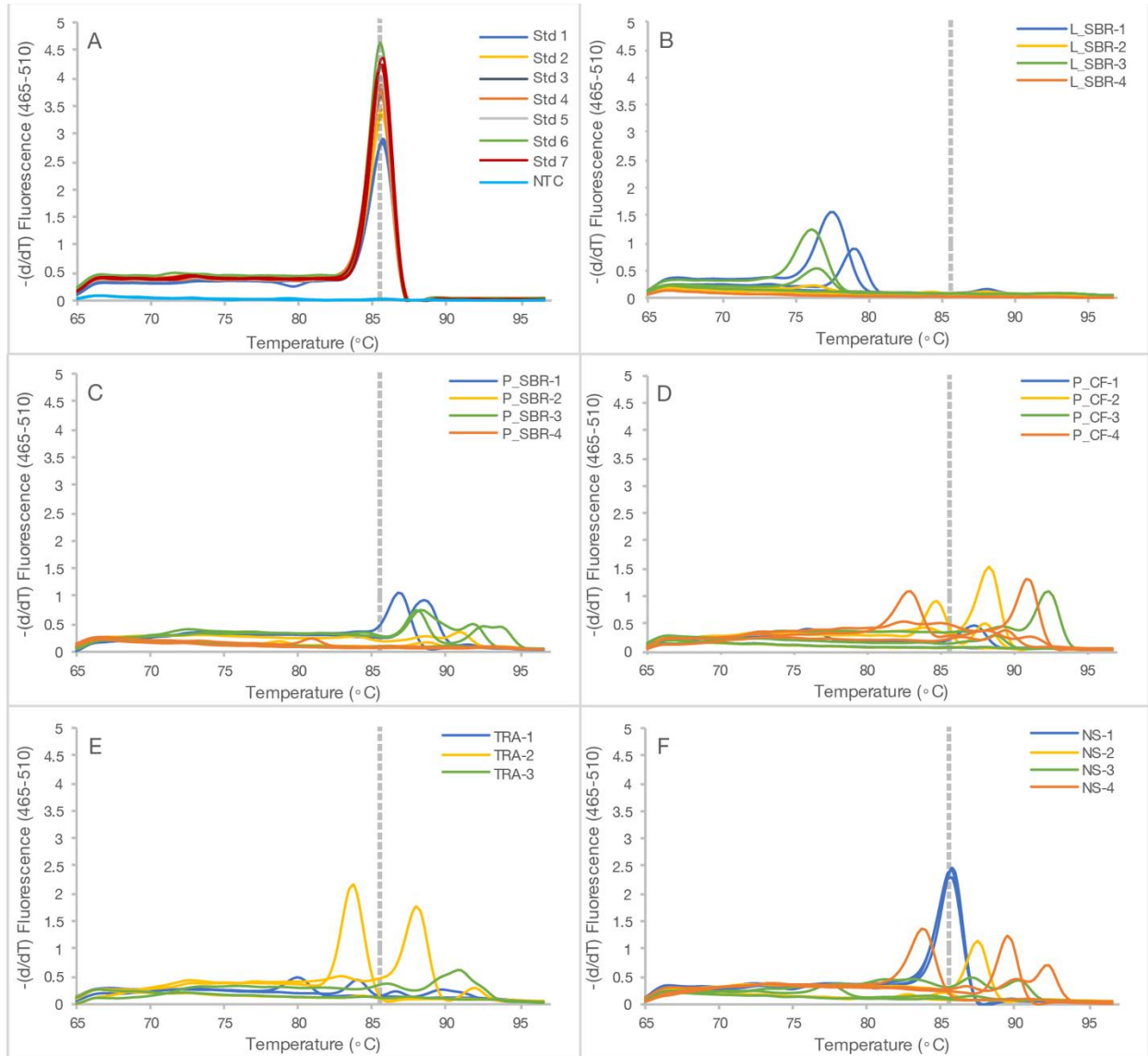


Figure S3-12. Comammox *Ca. N. inopinata* assay (Inopinata amoA-410F/815R) melting peaks after 45 cycles for A) *N. inopinata* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperatures for the standard. Triplicate data series are shown for each sample.

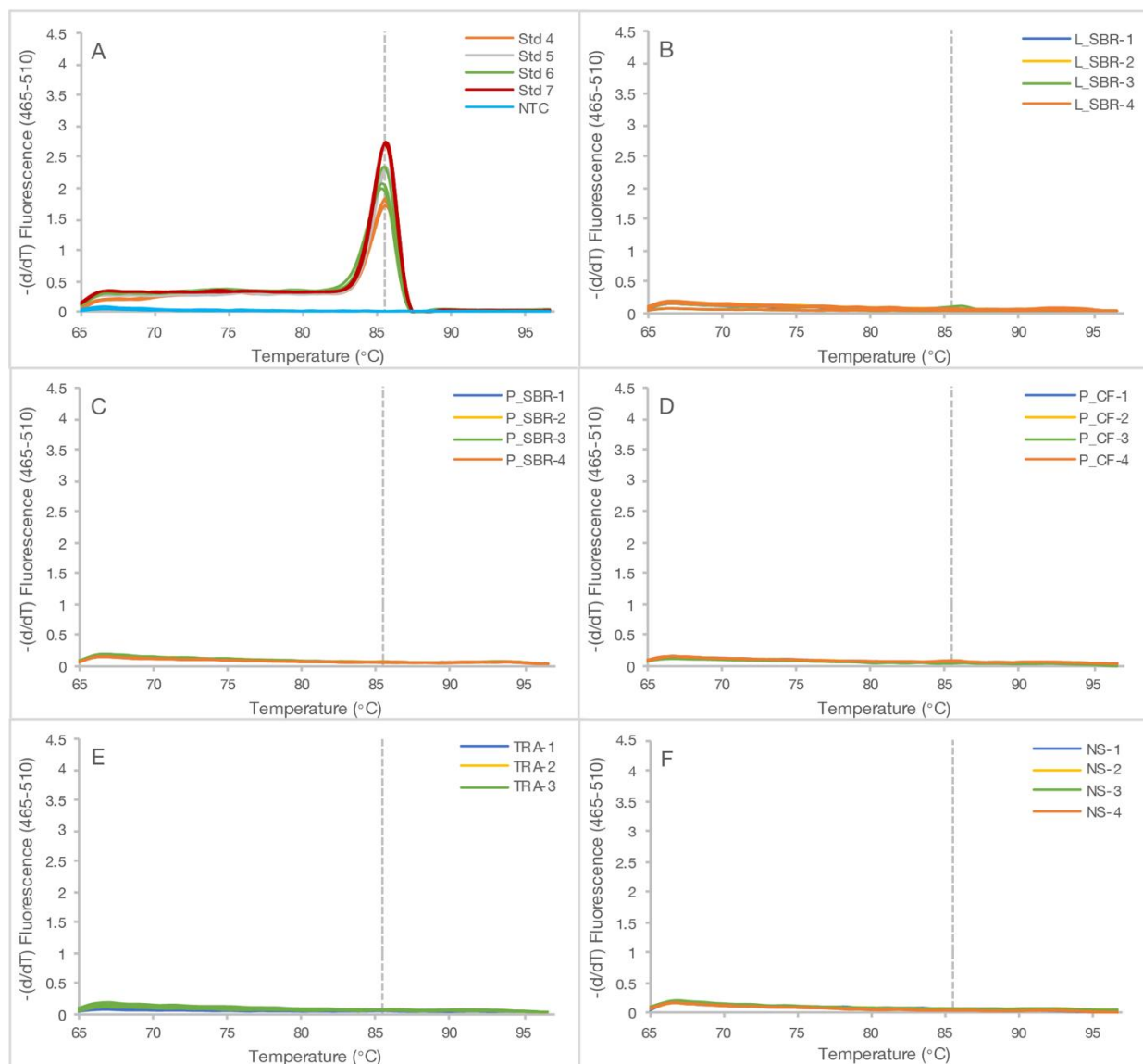


Figure S3-13. Comammox *Ca. N. inopinata* assay (Inopinata amoA-410F/815R) melting peaks after 30 cycles for A) *N. inopinata* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample.

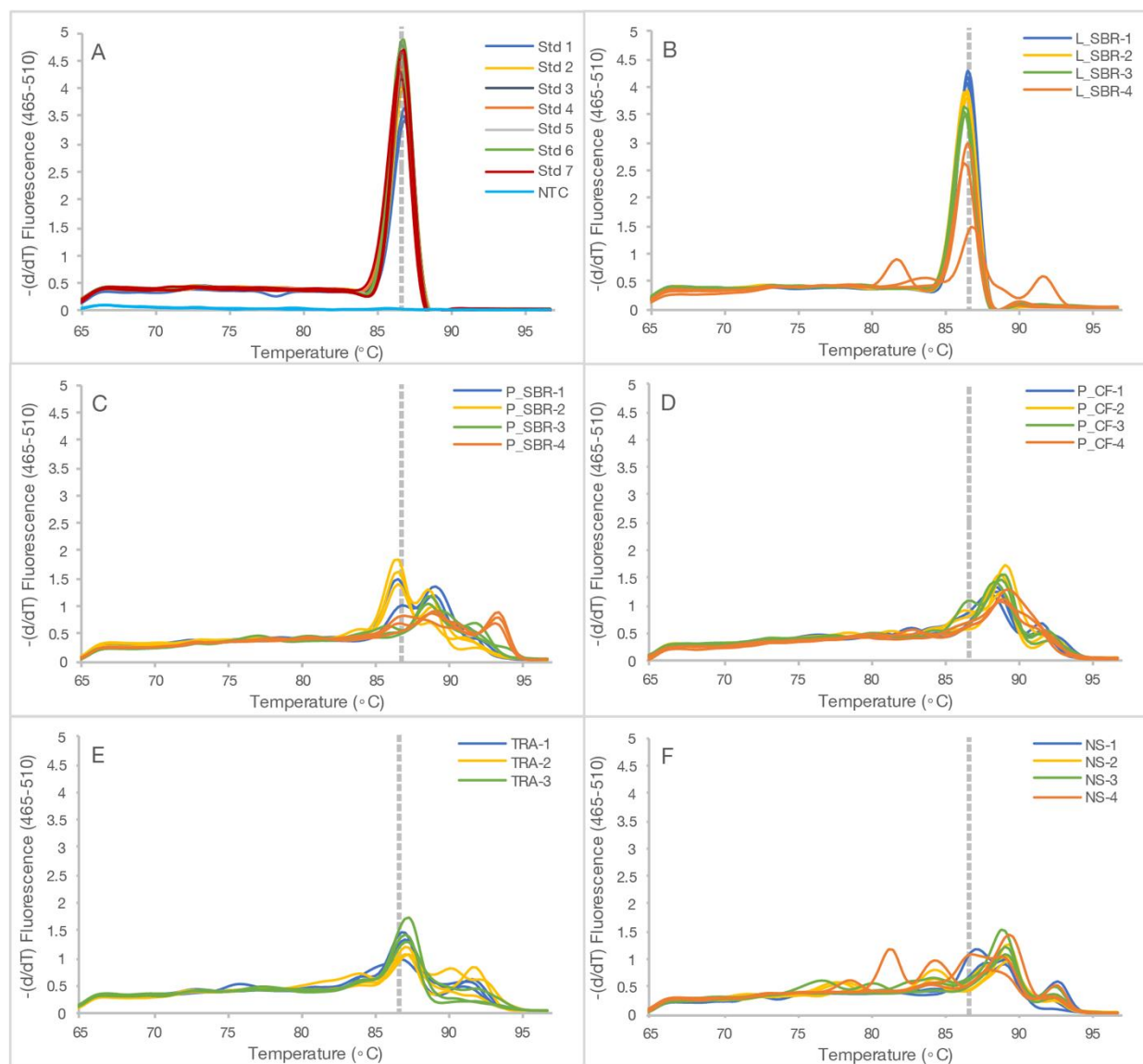


Figure S3-14. Comammox *N. nitrificans* assay (Nitrificans *amoA*-463F/836R) melting peaks after 45 cycles for A) *N. nitrificans amoA* standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperatures for the standard. Triplicate data series are shown for each sample.

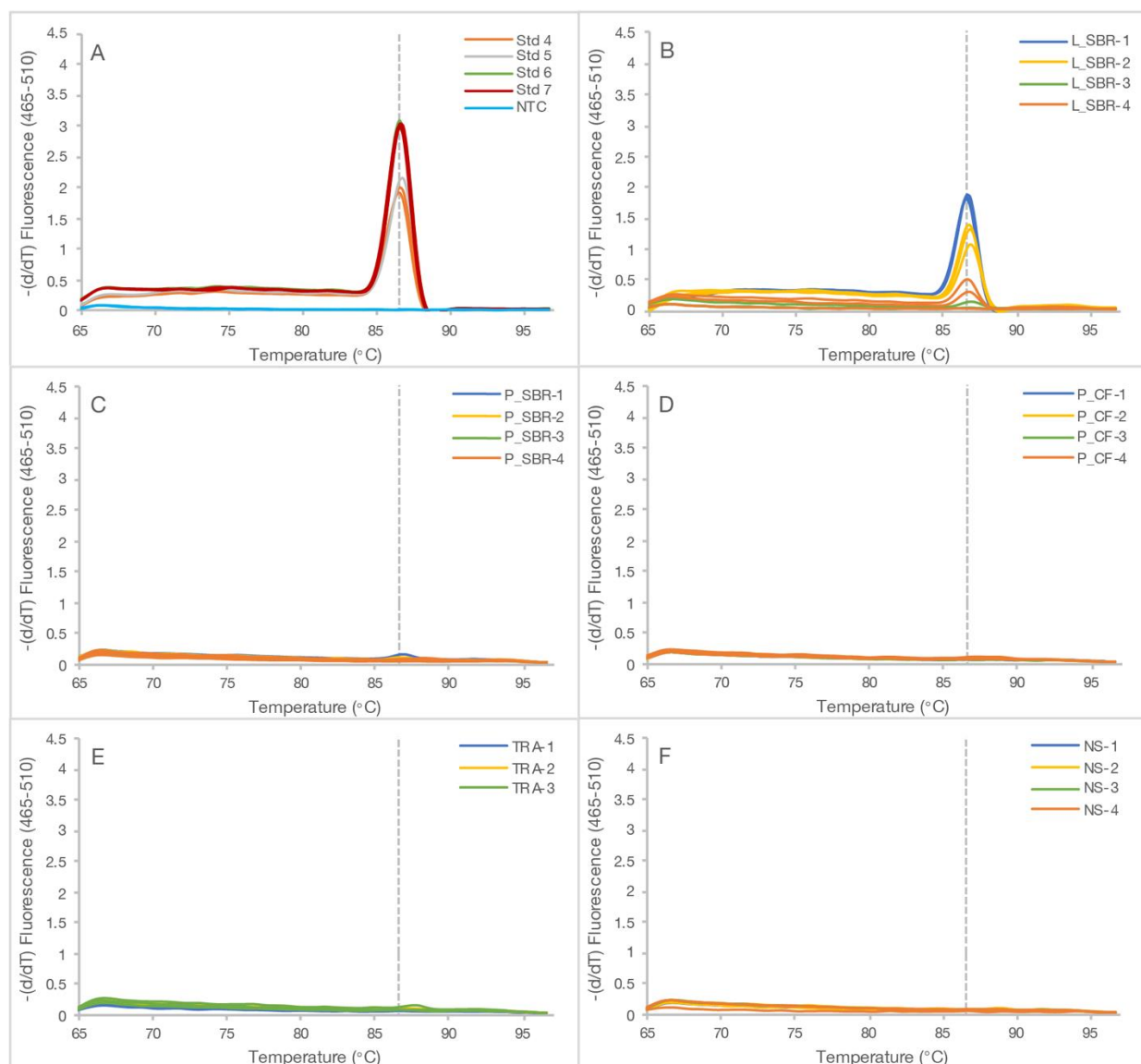


Figure S3-15. Comammox *N. nitrificans* assay (Nitrificans amoA-463F/836R) melting peaks after 30 cycles for A) *N. nitrificans* amoA standard curve, B) L-SBR samples, C) P-SBR samples, D) P-CF samples, E) TRA samples, and F) NS samples. The dashed lines in all panels represent the melting temperature for the standard. Triplicate data series are shown for each sample.

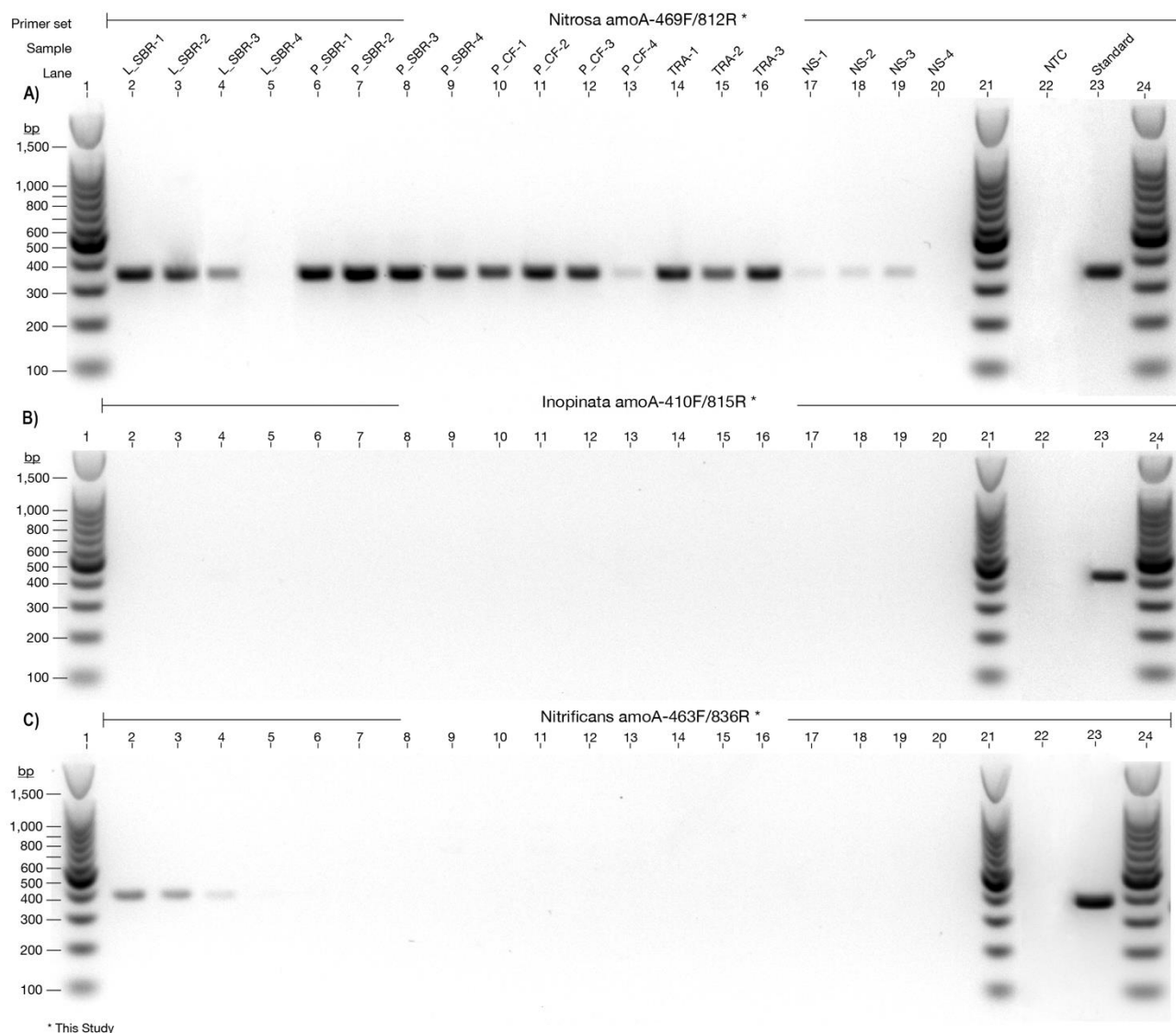


Figure S3-16. A qPCR assay with 30 cycles was generated to test each new primer set targeting A) *Ca. N. nitrosa*, B) *N. inopinata*, and C) *N. nitrificans* in one DNA sample from each bioreactor (lanes 2-20) with its corresponding non-template control (NTC, lane 22) and standard (lane 23). The expected band was present with *amoA* template from each external standard and some bioreactor samples. No off-target products were observed in the non-template control (NTC), bioreactor samples, or standards. Lanes 1, 21, and 24 correspond to 100 bp DNA size marker.

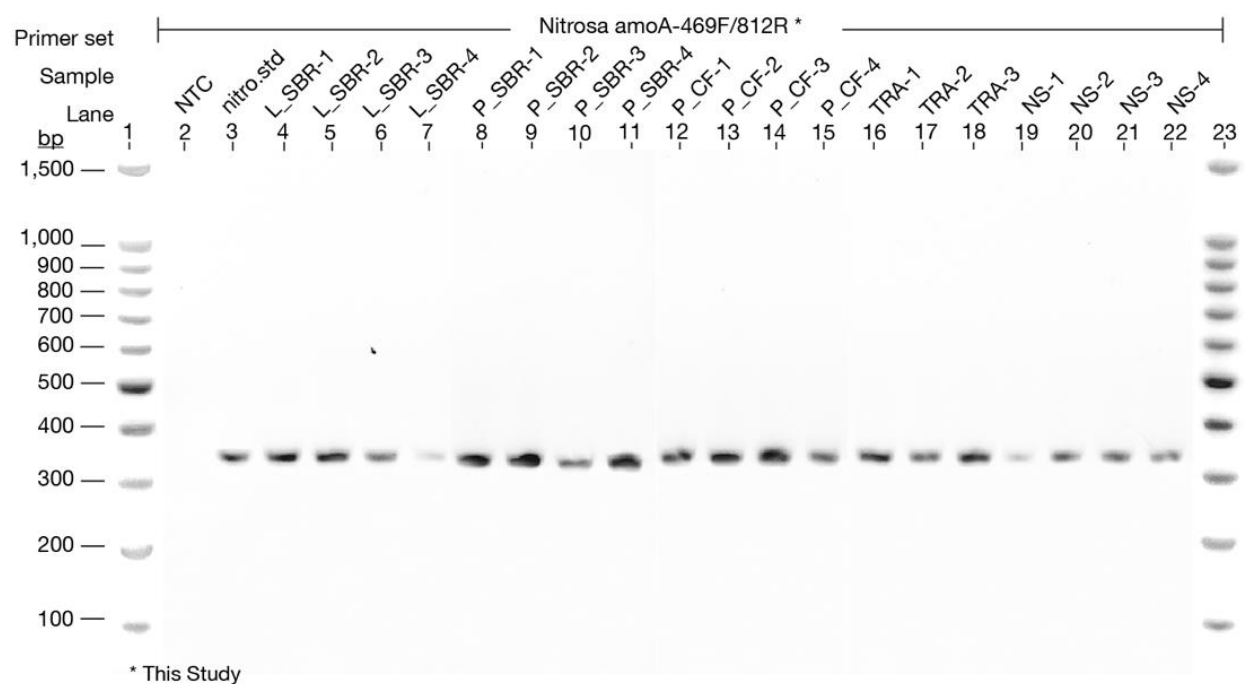


Figure S3-17. Agarose gel electrophoresis of qPCR products amplified after 45 cycles with the primers-set designed in this study targeting the *amoA* gene of *N. nitrosa*. The first and last lanes correspond to 100 bp DNA size marker.

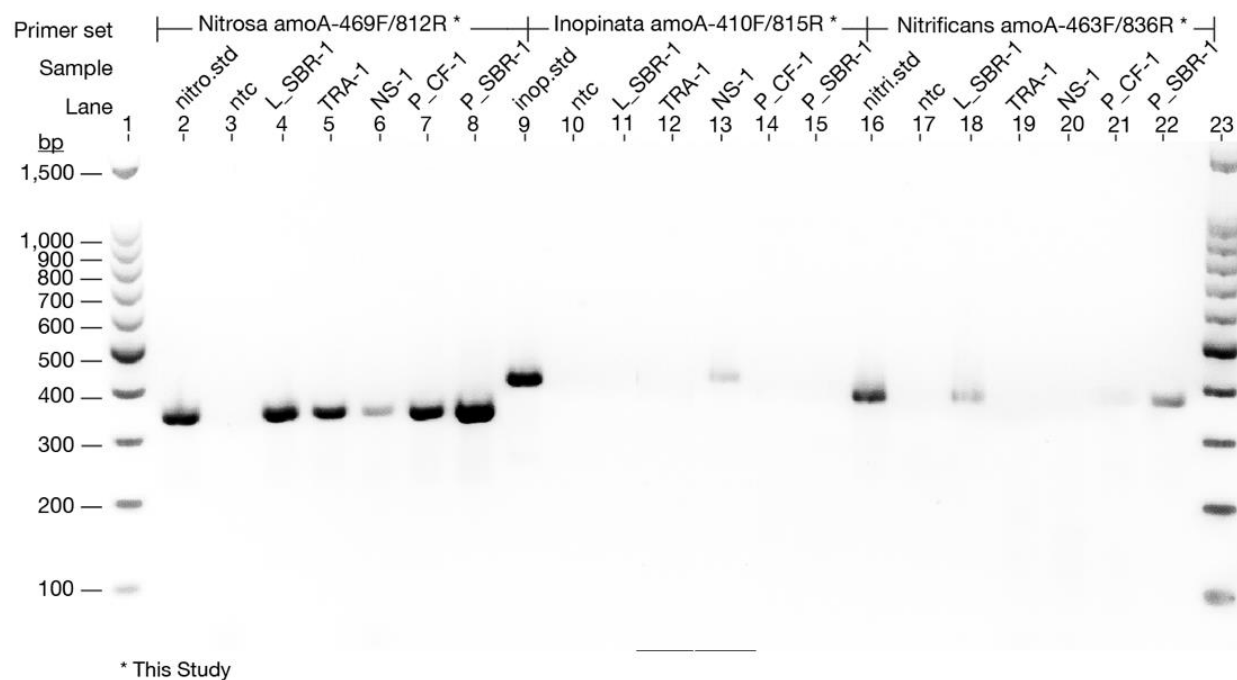


Figure S3-18. A PCR assay was generated to test each new primer set targeting *Ca. N. nitrosa*, *N. inopinata*, and *N. nitrificans* with its corresponding standard, a non-template control, and one DNA sample from each bioreactor. The expected band was present with *amoA* template from each external standard (*Ca. N. nitrosa* (nitro.), *Ca. N. inopinata* (inop.), and *Ca. N. nitrificans* (nitrif.)) and some bioreactor samples. No off-target products were observed in the non-template control (ntc), bioreactor samples, or standards. The first and last lanes correspond to 100 bp DNA size marker.

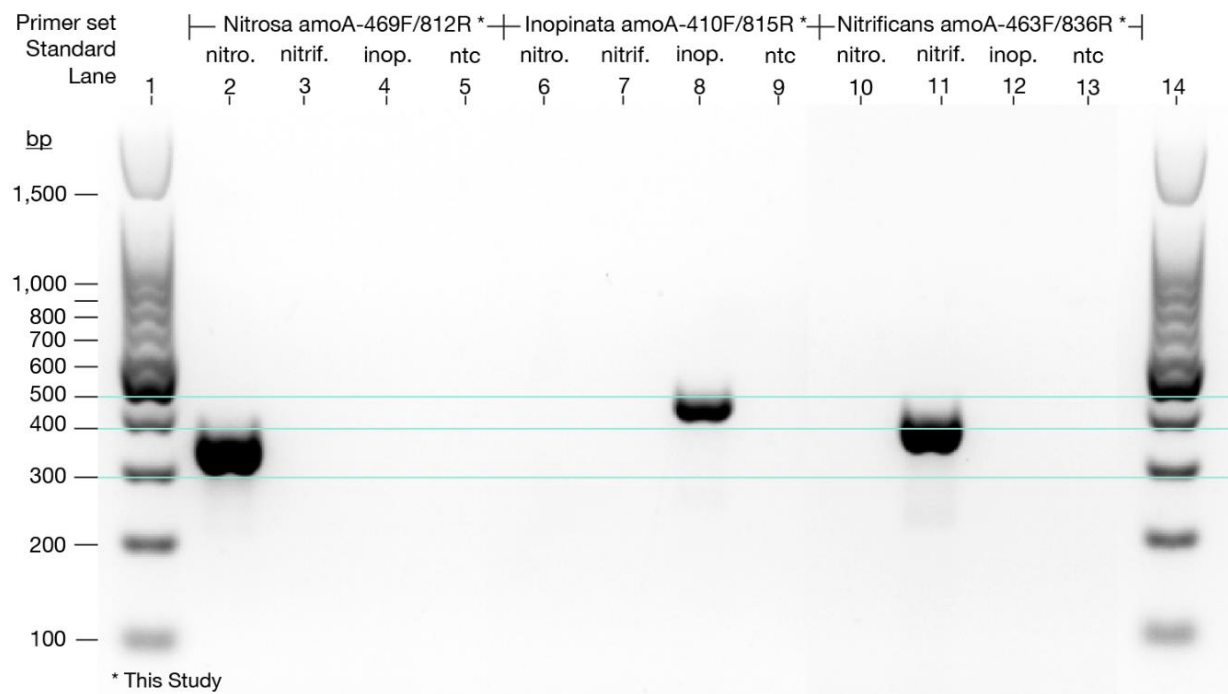


Figure S3-19. Cross-hybridization and primer dimers were not observed for the newly designed primer sets when combined in reaction with *amoA* template from each external standard (*Ca. N. nitrosa* (nitro.), *Ca. N. inopinata* (inop.), and *Ca. N. nitrificans* (nitrif.)) and non-template control (ntc). The first and last lanes correspond to 100 bp DNA size marker.

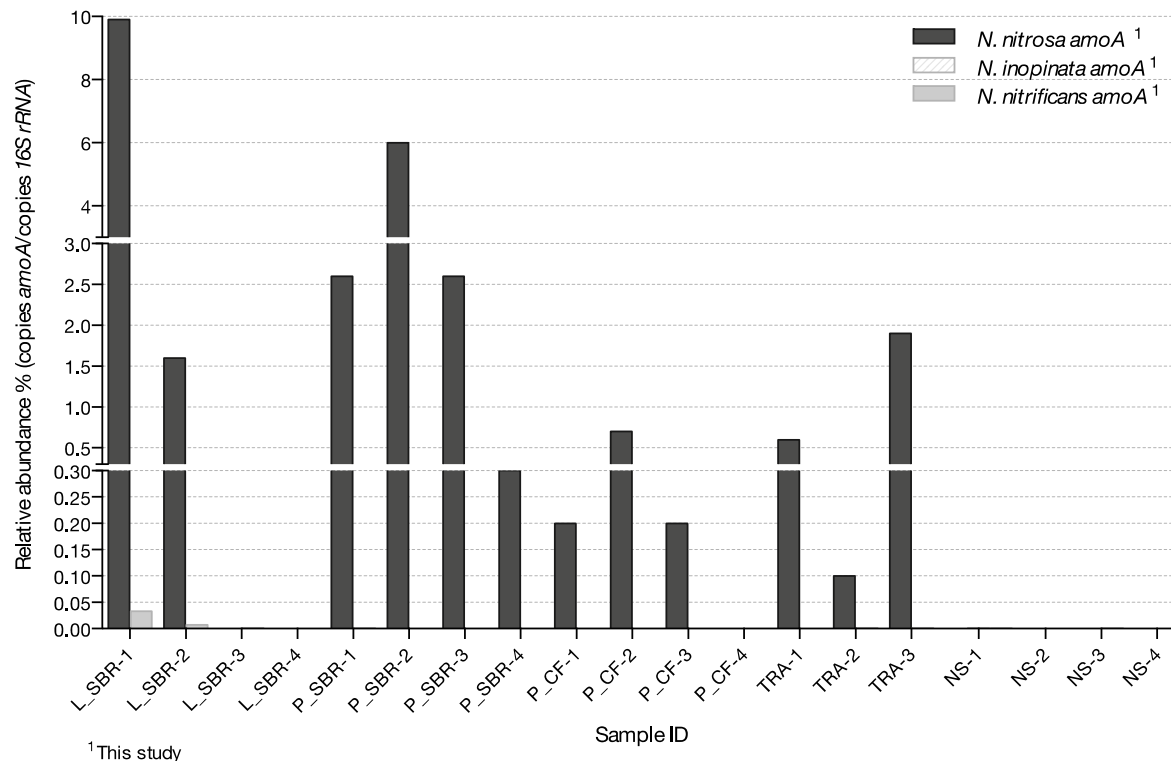


Figure S3-20. Percent relative abundance of comammox *Ca. Nitrospira amoA* genes to total bacteria (*amoA* gene copies/16S rRNA gene copies). Note the y-axis is not linear and is separated into three sections to show relative abundances below 0.30%, between 0.30 and 3.0%, and between 3.0 to 10%.

Chapter 4

Global regulation by FNR on gene expression in a partial-nitritation anammox microbiome

This chapter is formatted as a manuscript to be submitted for publication with the following authors:

Beach, N.K., Myers, K.S., Seib, M., and D.R. Noguera

NKB, KSM, and DRN developed the research plan, methodology, and project goals. NKB operated the bioreactor, designed and performed laboratory work, as well as *in silico* metagenome and metatranscriptome analysis. KSM contribute to metatranscriptomic analysis and identification of FNR binding sites. NKB drafted the manuscript, tables and figures. NKB, KSM, MS, and DRN contributed to data analysis and manuscript writing.

Abstract

Many bacteria are facultative anaerobes, that is, they can grow and survive in the presence or absence of oxygen. In partial-nitrification anammox systems, we often find the same functional groups participating in the cycling of inorganic nitrogen for energy generation, either by aerobic or anaerobic respiration. How microorganisms respond to alternating microaerobic and anoxic environmental conditions may be explained at the molecular level by a Fumarate and Nitrate reductase Regulatory (FNR) protein, which is widely conserved across bacteria. In *Escherichia coli* (*E. coli*), transcription factors belonging to the FNR family sense oxygen, and activate and repress a wide variety of genes responsible for diverse functions— thus, deeming FNR the title as a global transcriptional regulator. In particular, FNR is often responsible for regulating the switch between aerobic and anoxic dissimilatory metabolism; which contributes to the cycling of inorganic nitrogen in the partial-nitrification anammox microbiome. Since measuring transcriptome-wide gene-expression can reveal regulatory mechanisms that control how bacteria respond to changing environments— we employed a metagenomic and metatranscriptomic approach to explore the potential influence of FNR on the microorganisms within the partial-nitrification and anammox microbiome. We extracted and assembled highly complete, quality metagenomic assembled genomes (MAGs) and identified four of the most transcriptionally active microorganisms in the bioreactor as strains of the anammox *Candidatus* (*Ca.*) *Brocadia fulgida*, the ammonia oxidizing bacteria *Nitrosomonas europaea*, as well as a nitrite oxidizing bacteria belonging to the *Nitrospira* genus and a heterotrophic denitrifying bacteria belonging to the Chlorobi phylum. Each of these MAGs were highly identical to those recently published from two

other distinct PNA systems, one in Utah and the other across the globe in the Netherlands. We then used a published and experimentally refined FNR position weight matrix (PWM) to predict FNR binding sites across each MAG and we clustered gene expression profiles to identify a subset of oxygen-responsive changes in transcript abundance. The transcriptome-wide measurement of gene expression dynamics combined with the *in silico* FNR prediction provides evidence for global FNR regulation of the switch between aerobic and anaerobic energy metabolism in these four distinct microorganisms.

Introduction

Bacteria are found in diverse ecosystems, in which specific environmental conditions experience cyclical variations (e.g., day/night, summer/winter) and are sometimes unpredictable with ever-changing resource availability. The way bacteria continually adapt to changing environments is often controlled at the transcriptional level, allowing it to activate alternative metabolic pathways for energy generation and resource balance [1]. In other words, bacteria utilize a variety of regulatory proteins that promote or inhibit gene transcription, ‘turning on’ or ‘turning off’ certain genes as they are needed. One particular regulatory protein, the Fumarate and Nitrate reductase Regulatory protein (FNR), is among the most widely studied global regulators, activating and repressing transcription of a wide variety of functional genes. FNR is widely conserved across bacteria and is often required for the shift between aerobic and anaerobic metabolism [2-5]. Most of our knowledge concerning FNR and how it regulates gene expression has been gained with experiments on pure *Escherichia coli* (*E. coli*) cultures [6]; yet, only a fraction of microorganisms on earth have been identified and cultured [7]. *In vivo* approaches for mapping FNR binding sites across a particular genome remain extremely difficult and expensive, or impossible within microbiomes.

As we aim to improve understanding of nitrogen (N)-cycling microorganisms, we take an exploratory *in silico* approach to predict and analyze FNR-targeted gene regulation at the microbiome level. We take a unique opportunity to explore the adaptation of a partial-nitritation anammox (PNA) microbiome to repetitive and rapidly changing microaerobic and anoxic environment. Processes utilizing anammox bacteria for N removal have become extremely popular toward energy-saving wastewater treatment

[8-14]—and for good reason. Anammox organisms do not require oxygen or organic carbon as they convert ammonium (NH_4^+) and nitrite (NO_2^-) directly to nitrogen gas (N_2) [8]. The PNA process is viewed as a ‘shortcut’ to the classic biogeochemical nitrogen cycle, and has been frequently applied to the reject water from sidestream wastewater treatment processes, since anammox bacteria are not only sensitive to low temperature and high oxygen but also require a long solids retention time due to their slow growth rate [15]. Many of the most abundant and active functional groups in PNA systems are ubiquitous— independent of the inoculum used, bioreactor configuration, and operation. Yet, we are just beginning to realize and appreciate the extent of metabolic versatility and interactions within these microbiomes [16-18]. With exploration comes understanding and improvements to biotechnology follow. Ultimately, our investigation revealed that distinct microbial populations sharing the same niche exhibit highly similar oxygen-dependent gene expression patterns, which could be explained by the global regulator FNR.

Results and Discussion

Characterization of reactor operation

For 3 years, we evaluated the performance of a laboratory-scale deammonification bioreactor (single-stage partial nitritation-anammox process), which treated the effluent from the Nine Springs Wastewater Treatment Plant struvite recovery process (Madison Metropolitan Sewerage District, Madison, WI, USA). At day 0, the reactor was inoculated using biomass from the full-scale deammonification process at the York River Treatment Plant (Hampton Roads Sewerage District, Seaford, VA, USA). After an initial

period of poor performance, and observed loss of total N removal, reactor performance appeared to stabilize to ~90% total nitrogen removal by day 336. A representative nutrient profile after stable operation is shown in Fig. 4-1A.

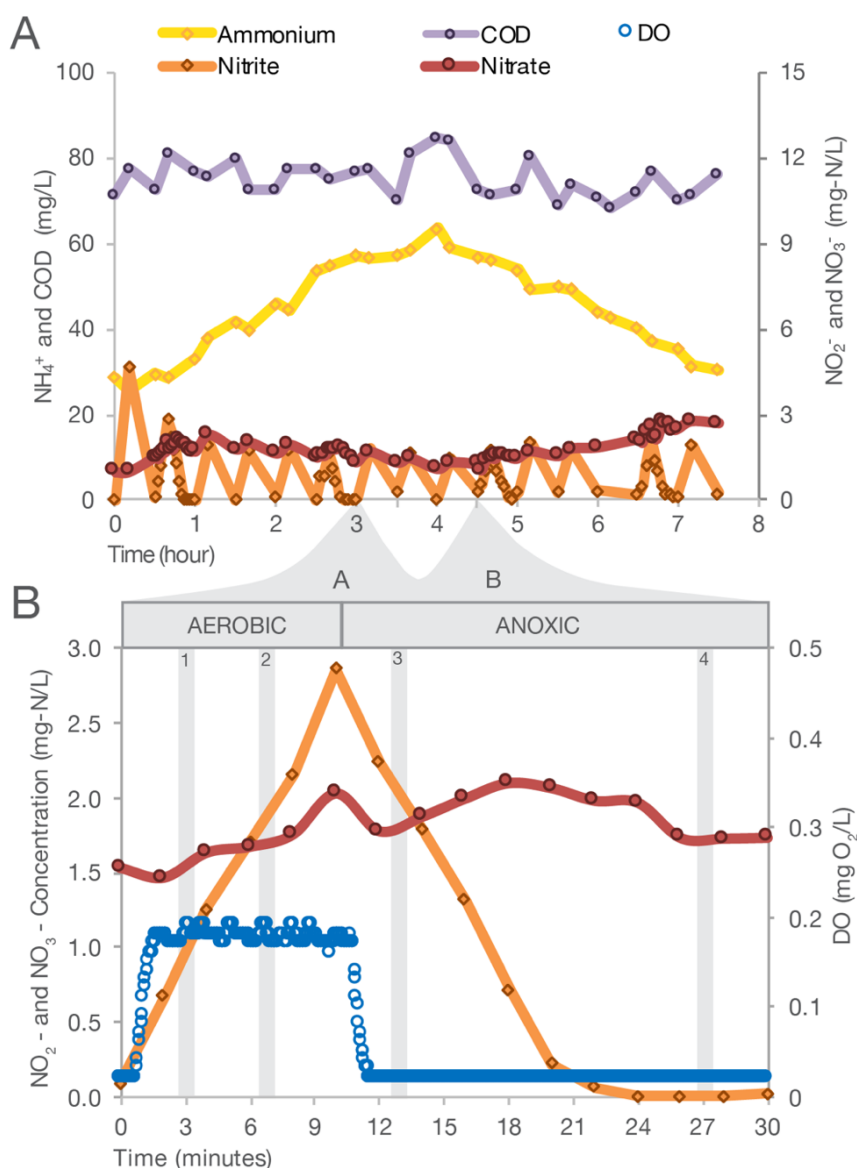


Figure 4-1. Bioreactor performance shown through: (A) A typical chemical profile during a single 8-hour cycle of operation; the plotted COD represents is the soluble fraction; and (B) a closeup of a 30 min time interval, defined by one aerobic and one anoxic cycle. RNA sampling were collected during time interval A and B, between 3-3.5 h and 4.5-5 h, respectively. In each interval, samples were collected during the aerobic period at 3 min and 7 min into the interval, indicated by '1' and '2' shaded regions. Additional samples were collected during the anoxic period at 13 min and 27 min, indicated by '3' and '4' shaded regions.

In a typical 8-h cycle, ammonium accumulated slightly in the reactor during the first four hours (Fig. 4-1A), which is the time the reactor was slowly filled with reject water from the Nine Springs Wastewater treatment plant (Madison, WI) struvite recovery facility (Fig. 4-2).

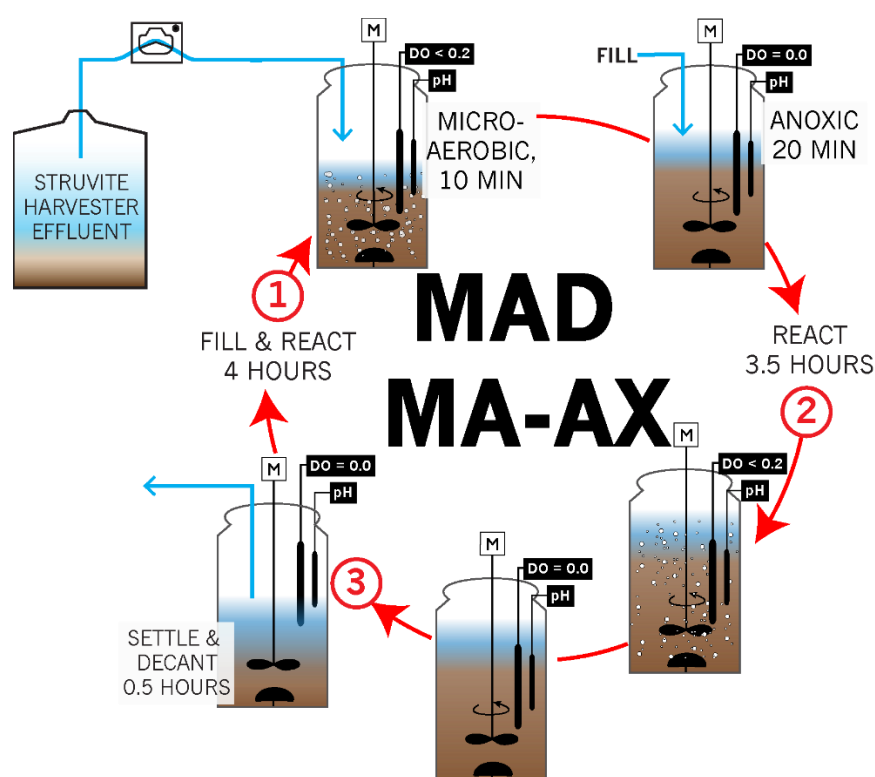


Figure 4-2. An 8-hour reactor cycle. (1) The system is slowly fed with the full-scale struvite harvester effluent during the first four hours of the cycle. During this time, conditions continuously cycle between a 10-minute micro-aerobic (< 0.2 mg O₂/L) and 20-minute anoxic period. (2) These intervals repeat for the next 3 ½ hours (between hours 4 - 7.5). (3) Then, the reactor is allowed to settle for 3 min, decants for 12 min and is idle for 15 additional min before the next cycle begins.

In the remaining 3.5-h, ammonium concentrations declined. In each of the 15 aerobic/anoxic intervals, nitrite accumulation was observed during the first 10-min in the aerobic phase and was subsequently removed in the 20-min anoxic period (Fig. 4-1B). During normal operation, dissolved oxygen concentrations fluctuated in the microaerobic period but remained below the control set point of 0.2 mg O₂/L (Fig 4-1B). Although pH and SRT were not controlled, pH within the reactor was an average of 7.5 ± 0.3 and total and volatile suspended solids (TSS/VSS) remained at an average of 1600 ± 550 mg/L and 1200 ± 440 mg/L, respectively (Table 1A).

Table 4-1. Reactor properties, chemical data statistics, and performance summary

Reactor properties, chemical data statistics, and performance summary											
A Reactor Properties											
	Units	Max	Min	Median	Average	Standard Deviation					
TSS	mg/L	3400	440	1500	1600	550					
VSS		3000	290	1100	1200	440					
pH	-	8.2	6.3	7.5	7.5	0.3					
B Chemical composition		Feed					Effluent				
	Units	Max	Min	Median	Average	Standard Deviation	Max	Min	Median	Average	Standard Deviation
Ammonium, NH ₄ ⁺	mg-N/L	370	45	190	200	60	290	0.10	31	50	58
Nitrate, NO ₃ ⁻		10	0	0.21	0.69	1.37	84	0.00	21	25	21
Nitrite, NO ₂ ⁻		81	0	0.03	2.51	10.5	170	0.0	0.1	4.4	18
Total COD (Unfiltered)	mg-O ₂ /L	2300	42	379	408	233	-				
Soluble COD (Filtered)		4800	64	330	260	350	500	7.0	67	76	38
BOD		2300	150	470	560	340	-				
pH	-	8.7	7.5	8.0	8.0	0.2	-				
n*	-	248					464				
C Performance summary											
	Units	Max	Min	Median	Average	Standard Deviation					
TIN removed	%	100	0.0	65	62	23					
Ammonium removed		100	1.0	85	76	25					
Nitrate produced		137	0.0	13	16	15					

* number of samples collected and included in statistics

Over the three years of operation, an average of $62 \pm 23\%$ of the total inorganic nitrogen (TIN) was removed from the feed, and an average of $16 \pm 15\%$ of the ammonium that was removed was converted to nitrate (Table 4-1B-C).

We were able to detect the common PNA community members at the genus level, such as *Candidatus (Ca.) Brocadia* and *Nitrosomonas* as well as *Nitrospira* using 16S rRNA tag amplicon sequencing. However, the average relative abundance of known anammox operational taxonomic units (OTUs) was much lower than what has been shown in other anammox systems (Persson et al., 2017; Suarez et al. 2017) (Supplementary Table S4-1). Relative abundance estimations using 16S rRNA have shown anammox belonging to the genus *Brocadia* in the range of 12-60% of the total community. In this study, *Ca. Brocadia* related OTUs had a total relative abundance of just $1.2 \pm 0.60\%$. Ammonia oxidizing bacteria (AOB) classified at the genus level as a *Nitrosomonas*, were $3.4 \pm 1.3\%$ and nitrite oxidizing bacteria, classified as *Nitrospira*, were detected at $1.0 \pm 0.6\%$ of the total community (Table S4-1). Thus, to help explore which microorganisms may be responsible for the nitrogen cycling and nitrogen losses observed in this system, we submitted samples for metagenomic sequencing from four select days of operation, day 77, 231, 350, and 454 that encompassed both low and high levels of TIN removal as well as varying relative abundance of each of the identified functional groups (Table S4-1). On day 454, we also collected triplicate biomass samples at 4 time points across the aerobic/anoxic intervals spanning between 3 - 3.5 h (interval A) and 4.5- 5 h (interval B) of a single reactor cycle (24 total samples) for RNA extraction and sequencing (Fig 4-1B) in order to identify active microbial populations in the system, gene expression patterns and coregulation.

Metagenomic sequencing, binning and quality

DNA from biomass samples on 4 select days of bioreactor operation (Table S4-1) were sequenced in duplicate, yielding eight separate metagenomic DNA read datasets. Of ~206 million metagenomic reads, ~ 201 million passed quality control criteria and were used as input for 1 metagenomic coassembly (Table S4-2). Coassembly resulted in 2,026,705 scaffolds, accounting for 91% of the quality filtered DNA reads, with N50 of 223,793 bp and an L50 of 1,480. The assembly also contained 83% of the quality filtered mRNA reads (Table S4-3), indicating that it also captured a large fraction of the metatranscriptome. With Anvi'o [19], we used a combination of automatic and manual binning to the co-assembly, which resulted in 77 manually curated, highly complete (>90%), low redundancy (<5%) metagenome-assembled genomes (MAGs) (Table S4-4). Together, 16 genomes accounted for 81% of the mRNA reads that mapped to the assembly, and therefore represent a major fraction of the active microbial community members present in the reactor on day 454 (Table 4-2; Table S4-4).

Table 4-2.Genome statistics of the top 16 MAGs recovered from the MAD MA-AX community. MAGs with the greatest relative expression are shaded green.

Genome statistics of the top 16 MAGs recovered from the MAD MA-AX community. MAGs with the greatest relative expression are shaded green.									
Bin #*	MAG ID	Phylum	Completeness (%)	Contamination (%)	Genome size (Mb)	# Scaffolds	N50 (scaffolds)	GC (%)	Predicted genes
Bin 58	CHB	Chlorobi	97	0.0	2.43	19	336,200	37.7	2,092
Bin 44	AMX	Planctomycetes	96	0.7	3.16	182	27,074	45.0	2,788
Bin 48	AOB	Proteobacteria	99	0.7	2.76	84	64,454	50.5	2,489
Bin 46	NOB	Nitrospirae	97	1.4	3.82	46	130,102	60.4	3,643
Bin 43	CFX1	Chloroflexi	97	1.4	3.80	262	26,194	57.0	3,582
Bin 57	PRO2	Proteobacteria	100	0.7	3.15	84	82,753	66.3	3,184
Bin 49	BCD2	Bacterioidetes	99	0.7	3.15	38	167,470	36.2	2,766
Bin 54R	PRO1	Proteobacteria	99	1.4	2.88	48	144,614	66.7	2,873
Bin 38	UNK1	Unknown Bacteria	98	1.4	7.65	49	347,644	58.8	5,664
Bin 68R	IGN1	Ignavibacteriae	97	1.4	3.45	19	447,731	42.3	2,679
Bin 50R	BCD1	Bacterioidetes	96	0.0	3.71	83	138,066	47.7	2,834
Bin 55	CFX2	Chloroflexi	98	1.4	4.06	62	273,379	55.2	3,681
Bin 56R	CFX3	Chloroflexi	97	2.2	3.79	18	577,897	53.6	3,515
Bin 32R	ACI	Acidobacteria	94	2.9	5.98	137	70,762	62.0	5,075
Bin 95R	UNK2	Unknown Bacteria	99	1.4	9.00	521	27,066	71.3	7,129
Bin 99	UNK3	Unknown Bacteria	96	0.0	3.64	43	213,845	48.6	3,032

* 'R' at the end of the Bin# indicates manual Anvi'o refinement to reduce redundancy

Phylogenetic analysis / Phylogenetic placement of MAGs

We compared the phylogenetic relationships of these top 16 recovered genomes with 84 published complete genomes and nearly complete metagenomes (Fig. 4-3) spanning across 7 different phyla. For this, we constructed a phylogenetic tree based on a concatenated protein sequences of 37 conserved bacterial single-copy marker genes [20]. When we analyzed the tree topology, we noted that several of our MAGs clustered closely with published genomes from other anammox systems. In particular, four of our MAGs (BCD1, BCD2, CHB and CFX2) grouped closely with genomes that were recently recovered from the full-scale Olburgen granular PNA reactor treating potato-processing wastewater in The Netherlands [17]. In addition, four other MAGs (AMX, CFX1, CFX3, and IGN) clustered closely with genomes recovered from a laboratory-scale anammox bioreactor located in the Goel laboratory at the University of Utah, which was initially inoculated with biomass enriched with anammox from the City College of New York [18]. Other notable groupings are AOB with the ammonia oxidizing bacteria (AOB) *Nitrosomonas europaea* (AL954747) and NOB with the nitrite oxidizing bacteria (NOB) ‘*Ca. Nitrospira defluvii*’ (NC014355)— both of which have been reported in wastewater treatment systems operated with low dissolved oxygen [21, 22].

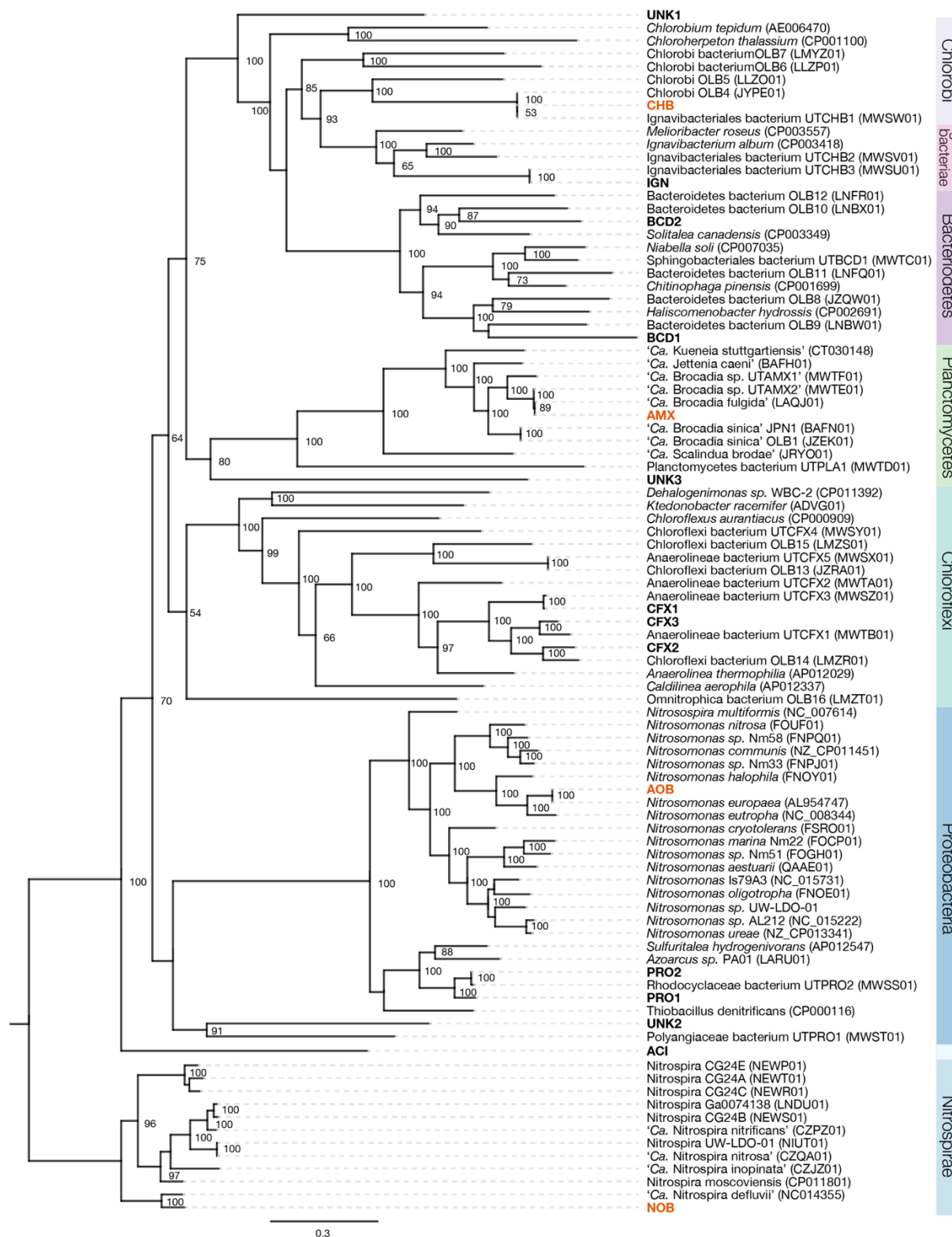


Figure 4-3. Phylogenetic tree generated from genome alignment of 16 most abundant MAGs recovered from the MAD MA-AX bioreactor. The tree includes MAGs recovered from this study (bold) and closely

related genomes downloaded from the NCBI genome repository. The four most active MAGs (based on metatranscriptomics) are shown in bold orange text. Accession numbers for each genome are provided in parentheses. Bootstrap values are shown at the nodes where the value was greater than 50. The tree was constructed using RAxML based on a set of 37 concatenated universal single-copy marker genes.

Microbial community abundance and gene expression

As a proxy for relative abundance and relative expression, we normalized the metagenomic read and mRNA read counts by the genome size and divided by the total DNA or mRNA normalized counts, respectively (Fig. 4-4). Surprisingly, gene expression did not always correlate with relative abundance in the MAD MA-AX bioreactor. Similar to previous attempts to determine relative abundance via 16S rRNA amplicon sequencing, genomes affiliated with *Nitrosomonas* (AOB), *Brocadia* (AMX), and *Nitrospira* (NOB) accounted for just 3.2%, 1.2%, and 0.6% relative abundance, respectively. Instead, genomes affiliated with Chlorobi (CHB), Chloroflexi (CFX1, CFX2, CFX3), Proteobacteria (PRO1, PRO2), and Bacteroidetes (BCD1, BCD2) encompassed ~89% of the MAD MA-AX community relative abundance. However, when evaluating relative expression, we had a much different observation with the genomes related to *Brocadia* (AMX), *Nitrosomonas* (AOB), and *Nitrospira* (NOB) as the second, third, and fourth most transcriptionally active microorganisms in the community, with relative expression of approximately 20%, 20%, and 3%, respectively. CHB1 dominated in both categories, with relative abundance and gene expression at ~55% and ~41%, respectively. Other organisms with less than 2% relative abundance and gene expression in the MAD MA-AX microbiome were affiliated with the phyla Ignavibacteriae (IGN), Acidobacteria (ACI) and other unknown Bacteria (UNK1, UNK2, UNK3) (Table 4-2; Fig. 4-3).

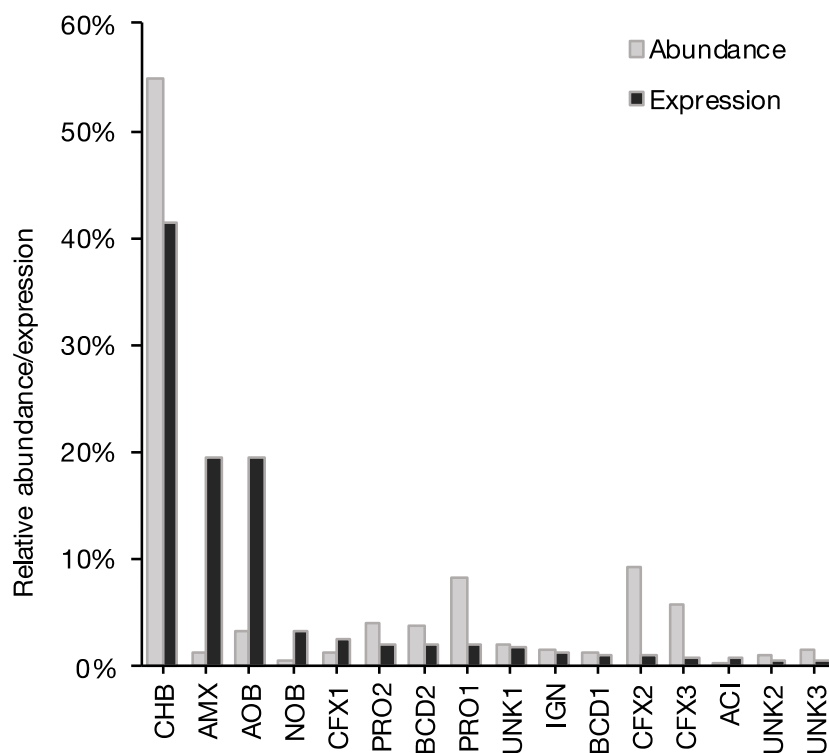


Figure 4-4. Relative abundance and gene expression of 16 most abundant metagenomic assembled genomes (MAGs) recovered from the MAD MA-AX bioreactor. Relative abundance was determined by the number of DNA reads mapped to each MAG, normalized by the corresponding genome size. Relative expression (transcript abundance) was determined by the number of cDNA reads that mapped to each MAG, normalized by the corresponding genome size.

The four most transcriptionally active MAGs

We now focus on the four most transcriptionally active MAGs, each of which are likely major contributors to the nitrogen cycling observed in the MAD MA-AX bioreactor, including: the potential heterotrophic denitrifier *Chlorobi* (CHB), ammonia oxidizing *Nitrosomonas* (AOB), nitrite oxidizing *Nitrospira* (NOB), and anammox *Brocadia* (AMX). We describe the assembled MAGs, identify relationships in phylogeny and nucleotide/amino acid identity, compare genomic content between the MAGs and published genomes, and identify inorganic nitrogen cycling genes and homologs in each

MAG. From the combination of metagenomic and time-series metatranscriptomic data, we examine the functional potential of each MAG, search for FNR transcription factor binding sites and explore global regulation of the genes involved in N transformation in the bioreactor.

To determine whether CHB, AOB, NOB, and AMX represented distinct species from closely related genomes, we performed average nucleotide identity (ANI) and average amino acid identity (AAI) comparisons with many of the genomes included in the phylogenetic analysis (Fig. 3), including: complete representative genomes from each phyla, candidate genomes, and metagenomes from the full-scale Olburgen anammox bioreactor (Olburgen, The Netherlands) and the laboratory-scale anammox bioreactor at the Goel Laboratory (University of Utah, USA) [17, 18]. ANI and AAI comparisons have been described as a robust method to compare genetic relatedness and define species-strain boundaries [23, 24].

Coding sequences that were predicted in each MAG were annotated/functionally profiled using Prokka [25], which initially predicts open reading frames with Prodigal [26] and then functional annotation with BLAST+ blastp [27] using all of the following databases in a hierarchical manner: (1) an optional custom database of annotated proteins, (2) all bacterial proteins in Uniprot [28], (3) proteins from complete genomes in RefSeq [29], and (4) hidden Markov model profile databases Pfam [30] and TIGRFAMs [31] as well as other databases for tRNA, rRNA, and tmRNA detection (as cited in the methods). If no matches were found among all databases, coding sequences were assigned the label 'hypothetical protein'. Since the publicly available databases commonly used to annotate the CDS of metagenomic assembled genomes do not

encompass the complete set of gene families involved in nitrogen cycling [32], we used Cd-hit to generate a custom non-redundant database of annotated proteins to include in the Prokka pipeline. This custom protein database was constructed with representative sequences from the manually curated and published 'NCycDB' database [32] combined with annotated protein sequences from complete and closely related genomes to the top four MAGs. We also used DIAMOND [33] to separately align proteins from our MAGs against those in 'NCycDB' [32], in order to obtain additional alignment information for the positive hits (i.e. e-value, percentage aligned, percent identity, bit score, etc.) and systematically identify N-cycling gene homologs based on the cutoff values discussed in the methods [34]. Additionally, we also performed searches with other publicly available orthology databases such as COG [35], eggNOG [36], and SEED subsystems [37]. For our genomes, COG assignments through the Anvi'o [19] workflow provided more gene assignments than either eggNOG or SEED. Specific details of annotations are outlined in each genome-specific section below.

We also analyzed four metapangenomes containing the MAGs along with several published genomes included in the phylogenetic analysis to compare, visualize, and characterize the unique and shared gene clusters across multiple microbial genomes using Anvi'o [19, 38]. Every genome was annotated with Clusters of Orthologous Groups (COGs) and single-copy bacterial and archaeal genes were identified using HMMER [39]. We binned gene clusters into categories based on their presence/absence to dissect gene cluster frequency and infer characteristics that may increase fitness in a cyclic low oxygen - anoxic environment. From this analysis, we were able to associate every predicted coding sequence (CDS) in each MAG, with a corresponding gene cluster,

category, and associated COG. Here, we investigated the gene clusters identified in 10 total genomes related to CHB, 17 total genomes related to AOB, 8 total genomes related to NOB, and 6 total genomes related to AMX. Specific results for each MAG are discussed further in the genome specific section below.

To infer potential functions and changes in gene expression across the microaerobic and anoxic interval we aligned mRNA transcripts from each sample collected during the microaerobic/anoxic interval to the predicted CDS for each MAG. The mRNA transcripts that aligned to genes in each MAG were counted and log fold change (LFC) in expression quantified with DESeq2 [40], which enables a more accurate quantitative analysis between time-series RNA-seq data by focusing on both presence and strength of differential expression, unlike commonly used approaches to normalize such as reads per kilobase million (RPKM) [41-43]. DESeq2 calculates the LFC and determines a false discovery rate (FDR) or probability for differential expression by modeling the count data for each gene using a generalized linear model (GLM) which strongly depends on a dispersion coefficient (critical for statistical interpretation of differential expression data) and is scaled by a sample-specific size factor determined via the median-of-ratios method which reduces the influence of outlier gene counts on the mean [44]. A ‘true’ dispersion (i.e. variability between replicates) is estimated for each gene by an empirical Bayes procedure where gene-by-gene specific maximum likelihood estimates for dispersion are calculated, a trend of dispersion over expression strength is determined (i.e. mean of normalized counts), and a ‘shrinkage’ parameter (i.e. how much the dispersion of a given gene varies from the consensus trend) is estimated and used to adjust the dispersion parameter [40]. A ‘true’ LFC in expression is calculated

for all genes by eliminating the influence of low read counts on LFC estimates (i.e. exaggerated LFC for genes with low counts), and again by determining ‘shrinkage’ via the empirical Bayes procedure, where ‘shrinkage’ is stronger for genes with low information, such as low counts, high dispersion, or low degrees of freedom [40]. These adjusted LFCs provides reliable interpretation of the strength of expression and can be used to rank genes based on dynamic changes in expression [40]. Finally, once each GLM is generated for each gene, the significance is tested via the Wald test and a p-value is determined and adjusted based on multiple testing to determine the FDR [45]. An $FDR < 0.10$ is recommended to define genes that are differentially expressed (90% confidence) [40]. In this study, however, due to the large number of genes to be considered between four genomes we use a more stringent $FDR < 0.05$ to define confident differentially expressed genes (DEG). This resulted in 1,169 DEGs in CHB; 2,059 DEGs in AOB; 1,997 DEGs in NOB and 1,972 DEGs in AMX.

To identify a potential regulon of the anaerobic transcription factor FNR in the community, we first verified that each MAG indeed contained a gene encoding for transcriptional regulators from the FNR family and that the genes were differentially expressed. Then, transcription factors with homology to FNR within the four genomes with the largest degree of transcriptomic coverage (CHB, AOB, NOB, and AMX) using an experimentally refined FNR position weight matrix (PWM) [4]. FNR binding sites were predicted 300 base pairs upstream of the translation start site of all genes in each MAG, assigned a site score, and filtered for a minimum score of 5 (i.e. the average site score for ‘true’ FNR binding sites in *Escherichia coli* [4]). From this workflow we identified: 1,038 total sites in AMX (37% of all genes), 449 total sites in NOB (12% of all genes), 757

total sites in AOB (30% of all genes), and 1,120 total sites in CHB1 (54% of all genes). The genes containing a predicted FNR binding site upstream were further reduced to include only those genes that were both differentially expressed ($\text{FDR} < 0.05$) and whose absolute LFC (sum of the absolute value of LFC over the time series) was greater than or equal to the median absolute LFC (i.e. the gene expression increased or decreased significantly), under the assumption that these genes are the most likely to be directly activated or repressed by FNR binding upstream. This resulted in a predicted FNR direct regulon in all the metagenomes with a total of 1,109 genes - with 302 genes in CHB (152 genes with increased expression during the anoxic phase and 150 genes with decreased expression), 306 genes in AOB (124 genes with increased expression during the anoxic phase of interval A and 182 genes with decreased expression), 136 genes in NOB (60 genes with increased expression during the anoxic phase interval A and 76 genes with decreased expression), and 360 genes in AMX (163 genes with increased expression during interval A the anoxic phase and 197 genes with decreased expression).

Since genes with similar function often respond in a similar manner to environmental stimuli, we clustered the time series expression of the predicted N-cycling gene homologs in each MAG to not only identify expression patterns but also to provide additional indication N transformation pathway activity [46]. Moreover, as bacterial adaptation to changing environmental conditions is often attributed to transcriptional regulation [5, 47], we explore if oxygen-dependent changes in transcript abundance were related to a co-localized predicted FNR binding site. In order to overcome common challenges experienced while clustering time series gene expression data (i.e. selecting number of clusters and modeling dependencies), we applied a specific nonparametric

model-based method (Dirichlet process Gaussian process mixture model [DPGP]) designed for this purpose [48], and were able to condense a total of 343 gene expression profiles into 10 groups with highly similar expression patterns (Fig. 4-7 and Fig. 4-18). This enabled us to not only identify co-expressed genes and support predicted gene homology but also explore if genes within a response group were potentially co-regulated as predicted by an FNR regulatory binding site.

Finally, we combined nucleotide and amino acid sequences for all genes, all annotations, pangenome gene cluster results, FNR site predictions, and differential expression data into a single database for each MAG (Supplementary Data Files 1-4) so that we could directly compare functional assignments, expression, FNR site scores, and for the purposes of this study -examine genes involved in inorganic nitrogen cycling and aerobic respiration .

Ultimately, we find the similarity between the organisms in this bioreactor and those of the full-scale Olburgen bioreactor striking due to differences in biomass composition (granular sludge versus flocculant biomass), wastewater sources (industrial versus domestic wastewater), geographic location (USA versus the Netherlands), and in some operational parameters such as HRT (5h versus 24h) and pH control [49]. Nevertheless, the two systems share several similarities, such as: both operate as a CANON-type process, single-stage bioreactor with intermittent aeration, and interestingly, the Olburgen bioreactor is also preceded by a struvite harvesting system [50]. In addition, despite differences in the wastewater source, influent chemical composition to the two reactors is very similar, with ammonium and COD concentrations to the Olburgen

reactor between approximately 200-320 mg $\text{NH}_4^+\text{-N/L}$ and 150-225 mg COD/L , respectively [17].

The CHB metagenome

The assembled CHB metagenome was 2.43Mbp with GC content of 37.7% and based on presence of single copy genes was determined as 97% complete and 0% redundant (Table 4-2). CHB had a comparable GC content to both *Chlorobi* bacterium OLB4 [17] and *Ignavibacteriales* bacterium UTCHB1 [18], which were 37.6% and 37.5%, respectively. CHB contained 2,092 predicted coding sequences with 707 not identified and labeled as hypothetical proteins (of which nearly half were assigned a COG category). CHB also had 102 CDS that were predicted N-cycling gene homologs (~5%), summarized in Fig. 4-5. Overall, 1,682 of CHB CDS were assigned to COG categories (80%) and 848 were assigned to SEED subsystems (41%).

From the phylogenetic tree, it was unclear whether the CHB genome may be related to *Ignavibacteriae* phyla (Fig. 4-3) or if it belonged instead to the *Chlorobi* phyla; however, when we calculated the ANI and AAI comparison between CHB and the *Ignavibacterium album*, they were both well below the species-strain definition threshold of 95% (68.9% ANI and 44.5% AAI) [23, 51, 52]. When we compared CHB with 9 other published genomes (Fig. 4-6), we observed that CHB was nearly identical to *Chlorobi* bacterium OLB4 (accession number JYPE01; referred to from here on as OLB4) and *Ignavibacteriales* bacterium UTCHB1 (accession number MWSW01; referred to from here on as UTCHB1) (ANI 99.7% and 99.7%, respectively; fraction aligned 97.5% and 96.6%, respectively) [17, 18] (Fig. 4-6). None of the other ANI values with the CHB genome

were greater than 75%, supporting the classification as a Chlorobi bacterium. Moreover, the AAI of CHB with OLB4 and UTCHB1 was 99.5% and 99.6%, respectively, indicating the three organisms likely belong to the same species (Fig S4-1 and S4-2).



Figure 4-5. Inventory of genes encoding for enzymes involved in aerobic respiration, denitrification, and transcriptional regulation in the four most active MAGs. Genes that were present in a genome but were not differentially expressed ($FDR > 0.05$) were assigned the minimum absolute dynamic LFC value of 0.1. Genes that were not present in a particular MAG are indicated by the white squares. The dynamic changes in gene expression are shown as a gradient from light blue to dark purple, between minimum and maximum dynamic log₂ fold change in expression (~0.1 - 12) (i.e. the sum of absolute value of log₂ fold change over the entire time series).

Since the shared gene content between genomes also provide effective projections of phylogenetic relationships [53, 54], we analyzed the distribution of gene clusters to further support the relationships among the genomes. Our pangenomic analysis of the 9 neighbor Chlorobi and Ignavibacteriae genomes from the phylogenetic tree (Fig. 4-3; Fig. 4-6) with a total of 25,925 gene calls resulted in 16,510 gene clusters. CHB genes were grouped into 2,076 clusters, which was very similar to OLB4 and UTCHB1 with 2,170 and 2,087 gene clusters, respectively. We grouped all gene clusters into 7 bins based on their occurrence across genomes: (1) Overall core gene clusters (n=105), (2) gene clusters shared exclusively between CHB and OLB4 (n=51), (3) shared gene clusters between CHB and UTCHB1 (n=8), (4) gene clusters shared between CHB, OLB4, and UTCHB1 (n= 1,261), (5) unique gene clusters to CHB (n= 33) (6) other singletons (i.e., gene clusters associated with a single genome; n=12,677), (7) other gene clusters that do not fit any of these classes (n=2,375), (Fig. 6). The singletons and CHB, OLB4, and UTCHB1 core gene clusters corresponded to ~50% and ~5% of all clusters, respectively.

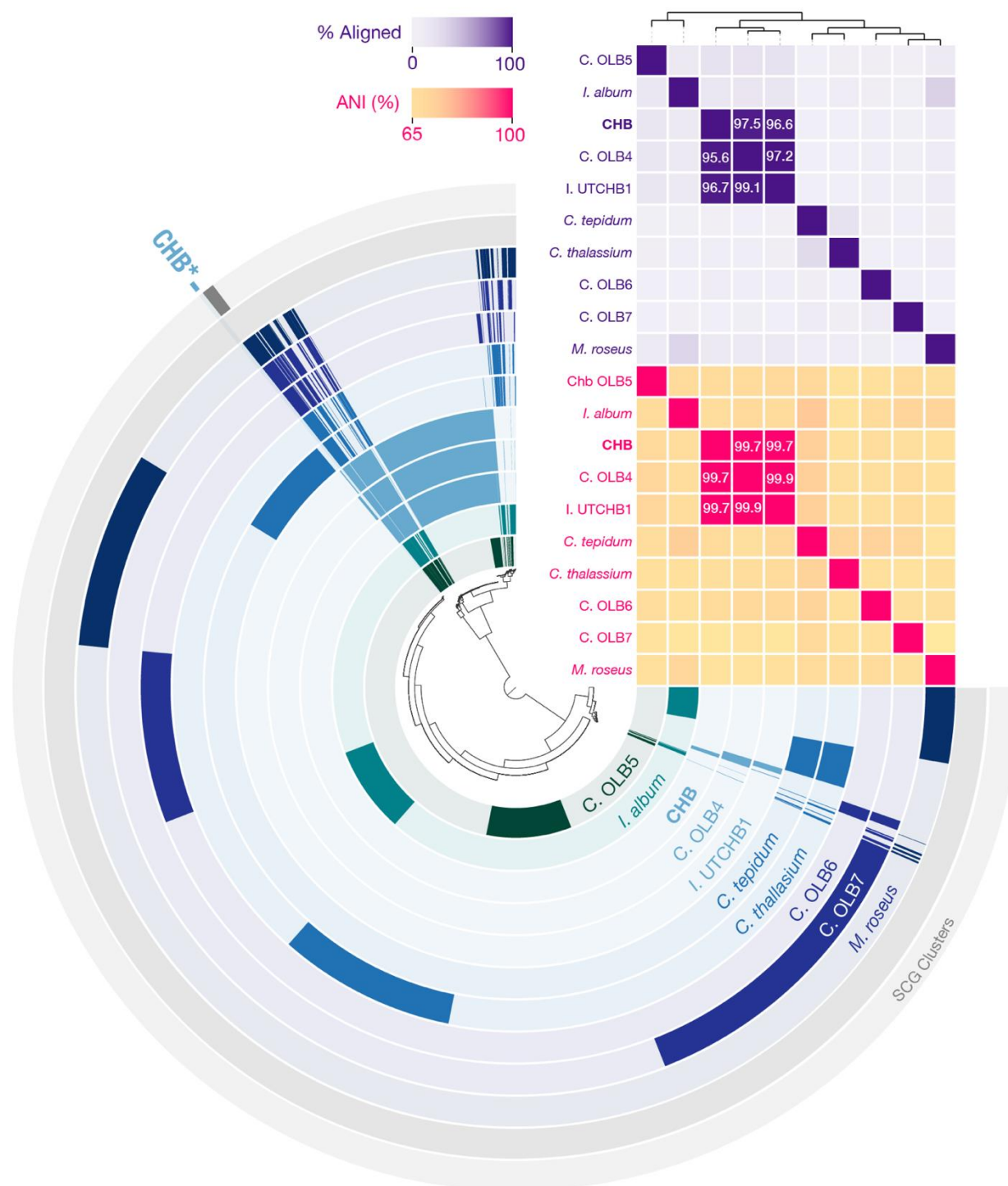


Figure 4-6. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the CHB metagenome (this study) and 9 published Chlorobi and Ignavibacteriae genomes from the phylogenetic tree. All genomes were greater than 90% complete with less than 2.5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 16,510

GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: 'Core' (gene clusters shared between all genomes), 'Shared' (gene clusters present in CHB and closely related genomes *Chlorobi bacterium* OLB4 [JYPE01] and *Ignavibacteriales* UTCHB1[MWSW01]), 'Unique' (gene clusters present in only CHB, indicated on the outside of all rings), 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices.

The CHB genome encoded for a complete denitrification pathway, with genes homologous to the membrane bound dissimilatory nitrate reductase genes (*narGHJI*) that reduce nitrate to nitrite, the nitrate transporter (*narK*), the copper-containing nitrite reductase (*nirK*) which reduces nitrite to nitric oxide, hybrid cluster protein (*hcp*; encoding for a nitric oxide reductase (NOR) type enzyme recently discovered in *E.coli* [55]), and nitrous oxide reductase (*nosZ*) [7] (Fig. 4-5). The median absolute LFC of all differentially expressed genes in CHB was 0.66, meaning that median gene expression was either increased or decreased by a multiplicative factor of $2^{0.66}$ or ~1.6. We use the median absolute LFC as a proxy for identifying and ranking dynamic genes (genes that are significantly turned on and off by the cell). The respiratory nitrate reductase operon (*narGHJI*, gene caller IDs 741-744) was actively transcribed by CHB, indicating its potential use in nitrate reduction in the bioreactor. The catalytic subunit, *narGH*, was the most dynamic of the nitrogen cycling genes with absolute LFC of approximately 5 in CHB and was up-regulated during the aerobic phases of both sampling intervals and subsequently down-regulated during both anoxic phases, which placed the expression pattern into Group 5 of the DEG clustering analysis (CHB 3 subgroup in cluster Group 5; Fig. 4-7; Supplementary Table S4-5).

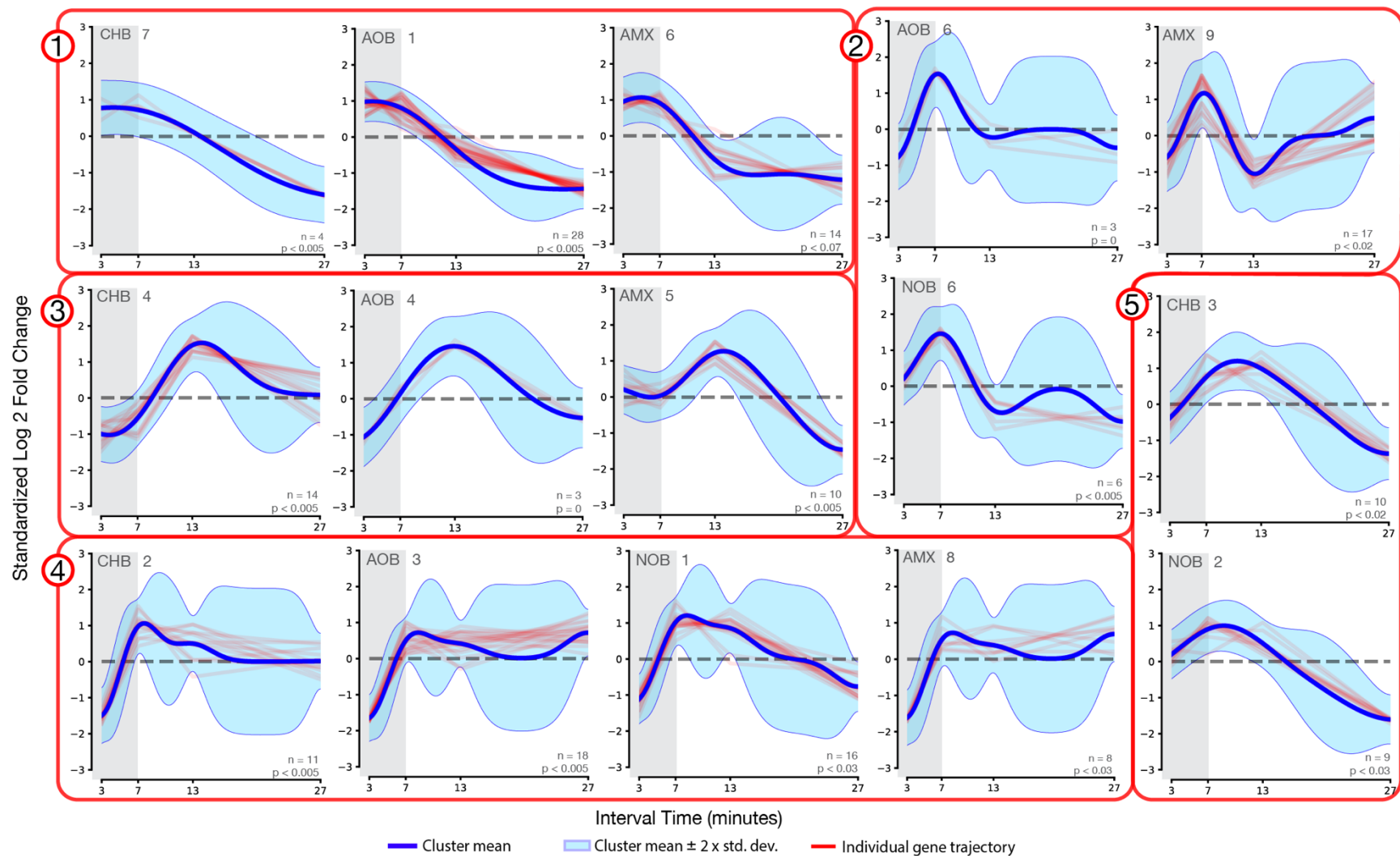


Figure 4-7. Response Groups 1-5 containing time-series gene expression profiles for the four most transcriptionally active MAGs in the MAD MA-AX microbiome. Expression profiles for differentially expressed genes related to inorganic nitrogen cycling were clustered with a nonparametric model-based method, DPGP, in each MAG. Gene expression response types for these MAGs were grouped based on similar response types. The y-

axis represents the standardized log₂ fold change in gene expression over the time series and the x-axis is the sampling interval time in minutes. The shaded region in each plot represents the samples collected during the microaerobic period, followed by an anoxic period. The upper left corner of each plot denotes the MAG-specific subgroup within a particular response group (numbers circled in red). The number of genes clustered and statistical significance are denoted in the bottom right corner of each individual plot

While an FNR site was detected ahead of *narJ*, it was not confidently detected upstream of the *nar* operon; thus, FNR may not be the primary regulatory mechanism for nitrate reduction with NAR. Nevertheless, the aerobic up-regulation in *nar* transcripts may indicate one way that the cell strategically prepares for the next cycle of anoxic conditions and anaerobic metabolism [56]. This nitrate reductase operon was also considered unique to CHB, OLB4, and UTCHB1 from the pangenome analysis.

Unlike the other two related genomes, we found several genes with predicted similarity to the copper-containing nitrite reductase (*nirK*) and the haem-containing nitrite reductase (*nirS*). We found 14 genes similar to *nirK* that were also differentially expressed (FDR < 0.05), following the guidelines provided by Pearson, 2013 [34] to infer which genes may be homologs of nitrite reductases (e-value < 10E-10, bit score > 50, percent identity > 30%, and coverage > 80% as cutoffs for gene homology). While the percent identity measurement can be helpful for determining evolutionary distance between genes, it is not considered reliable or sensitive for inferring homology [34]. Instead, the e-value and bits of similarity score (bit-score) are statistically significant measures of homology [34]. When determining gene homology, however, it is important to remember that although homologous sequences have similar structure and often similar function— the relationship between homology and function is not always predictable [34]. For this reason, we also analyze the gene expression patterns of all potential homologs to N-cycling genes and observe both with which genes they cluster and how they may be regulated during changes in environmental conditions.

Of the 14 potential *nirK* homologs, only 4 were actively expressed or repressed above the median absolute LFC with the most dynamic at 1.4 absolute LFC. This potential *nirK*

homolog (gene caller ID = 1844) was up-regulated during the aerobic phase of the interval B with expression remaining relatively constant during the anoxic conditions. This homolog clustered with gene expression patterns in Group 4 of the DEG clustering analysis (CHB 2; Group 4 in Fig. 4-7; Supplementary Table X). We also found five potential *nirS* homologs, two of which had transcription activity above the median absolute LFC (gene caller IDs 532 and 1445) with both following similar gene expression patterns to the *nirK*-1844 (gene name - gene caller ID) homolog mentioned above (CHB 2; Group 4 in Fig. 4-7; Supplementary Table S4-5). Interestingly, the one *nirS*-532 with the highest absolute LFC (~1.3) was also found in a shared gene cluster with OLB4 and UTCHB1, while these genomes were originally reported as not participating in nitrite reduction [17, 18]. While we did not detect an FNR site upstream of *nirK*-1844, we found a positive hit for an FNR motif upstream of both *nirS* homologs, thus FNR may be the oxygen responsive switch for nitrite reduction with *nirS* in CHB.

We found four genes similar to the hybrid cluster protein (*hcp*), which is a high affinity nitric oxide reductase type enzyme that is reported to reduce nitric oxide to nitrous oxide in *E. coli* [55]. Two of these potential *hcp* genes had differential change in gene expression above the median LFC, the most dynamic *hcp* gene (gene caller ID = 80; absolute LFC ~2.9) had gene expression dynamics similar to the *nar* genes, placing it into Group 5 (CHB 3, Group 5 in Fig. 4-7). This *hcp* gene homolog was also unique to CHB, OLB4 and UTCHB1 and may indicate that this particular species participates in nitric oxide reduction to nitrous oxide, perhaps for detoxification purposes [7, 55]. Like *nirS*, we also found positive hits for FNR upstream of *hcp*, and since the transcriptomic response of *hcp* to oxygen is similar to that of *narGHJI* but an FNR binding site was not

confidently detected upstream of the *narGHJ* operon—it remains unclear if both also share similar regulatory mechanisms.

Furthermore, we found that CHB also encoded for the nitrous oxide reductase (*nosZ*), reducing nitrous oxide to nitrogen gas, and is consistent with reports for the OLB4 and UTCHB1 genomes. Out of the 8 potential *nosZ* homologs, only 2 were dynamic above the median absolute LFC. The most dynamic of the potential *nosZ* genes also had the highest percent identity to a *nosZ* gene in the NCycDB database (100% id, 100% coverage; gene caller ID = 93; absolute LFC ~ 0.96) and was slightly up-regulated aerobically and steadily down-regulated anaerobically, placing the gene expression profile into Group 1 (CHB 7, Group 1 in Figure 4-7). The other potential *nosZ* gene (gene caller ID=1741; absolute LFC ~0.78) behaved slightly differently in the two sampling intervals. In interval A, *nosZ*-1741 gene expression dynamics followed similar expression patterns to *nirK*-1844 and both *nirS* genes mentioned previously (*nirS*-532 and *nirS*-1445), with a fast up-regulation between minutes 3 and 7 and slight down regulation anaerobically (CHB 2, Group 4 in Fig. 4-7) . In interval B, the *nosZ*-1741 gene was up-regulated at a slower rate between 3 and 13 minutes (CHB 4, Group 3 in Fig. 4-7). The only shared *nosZ* homolog between CHB, OLB4, and UTCHB1 (gene caller ID = 680,) was differentially expressed (FDR < 0.05) with different expression patterns between the two intervals but did not change above the median absolute LFC (LFC ~ 0.6). This *nosZ*-680 was subsequently placed into groups with virtually the opposite expression patterns: Group 6 (CHB 6, Group 6 in Fig. 4-8) for activity during interval A, and Group 9 (CHB 1, Group 9 in Fig. 4-8) for activity during interval B. Group 6 is characterized by a sharp aerobic down-regulation and slight anaerobic up-regulation. On the other

hand, Group 9 is characterized by a steady up-regulation over the entire time-series (Fig. 4-8). The difference in gene-expression between the two intervals is unclear; but may be due to a change in nitrous oxide production during either interval or due to changes in substrate availability since interval B occurred later in the bioreactor cycle. Here we found positive detection of an FNR binding site upstream of the most dynamic *nosZ*-93 and *nosZ*-1741, but not *nosZ*-680. The oxygen response of *nosZ* transcription is less clear than some of the other N-cycling genes in this MAG, however, the differences in *nosZ* oxygen response of these three homologs may indicate that they do not share the same nitrous oxide reducing function.

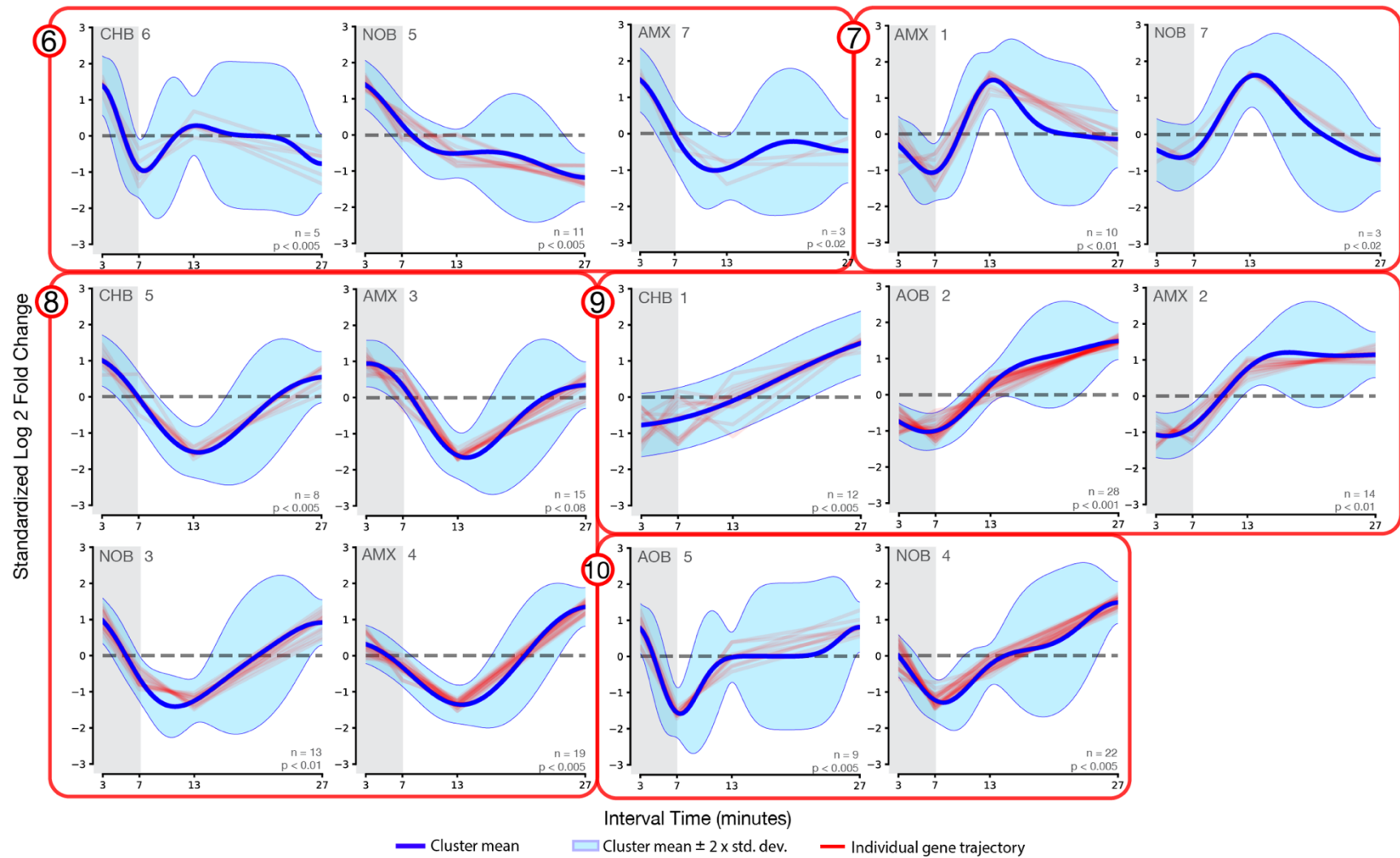


Figure 4-8. Response Groups 6-10 containing time-series gene expression profiles for the four most transcriptionally active MAGs in the MAD MA-AX microbiome. Expression profiles for differentially expressed genes related to inorganic nitrogen cycling were clustered with a nonparametric model-based method, DPGP, in each MAG. Gene expression response types for these MAGs were grouped based on similar response types. The y-

axis represents the standardized log₂ fold change in gene expression over the time series and the x-axis is the sampling interval time in minutes. The shaded region in each plot represents the samples collected during the microaerobic period, followed by an anoxic period. The upper left corner of each plot denotes the MAG-specific subgroup within a particular response group (numbers circled in red). The number of genes clustered and statistical significance are denoted in the bottom right corner of each individual plot.

The CHB genome also encoded for a cytochrome aa3-type terminal oxidase (*coxBA*) and a cbb3-type terminal oxidase (*ccoN*), both of which are characterized in detail within *Paracoccus denitrificans* [57]. These terminal oxidases are responsible for reducing oxygen to water in the last step of the electron transport chain of aerobic respiration. The aa3-type terminal oxidase is described as being highly efficient at generating a proton gradient across the cell membrane and is reported as advantageous for a cell under starvation conditions [58], but is also described as having low-affinity for oxygen [18]. The cbb3-type terminal oxidase is described as having high-affinity for oxygen, is expressed under microaerobic conditions, and is preceded by a promoter that contains an FNR binding site [57]. The gene clusters encoding both aerobic respiratory enzymes was shared uniquely between CHB, OLB4, and UTCHB1, suggesting that all three of these organisms are facultative anaerobes, utilizing oxygen for respiration when it is available. The discovery of an aa3-type oxidase in UTCHB1 is consistent with the original report [18], and it was not reported whether OLB4 was involved in any aerobic respiration [17].

As expected, the genes encoding *coxA* and *coxB* (gene caller IDs of 1883 and 1882, respectively) were co-expressed with a mean absolute LFC of 1.2 and were clustered with gene response Group 9 (CHB 1, Group 9 in Figure 4-8). Group 9 is characterized by an overall, steady up-regulation primarily occurring during the anaerobic phase. The gene encoding *ccoN* (gene caller ID 1697; absolute LFC ~1.13) followed the same expression pattern as *coxBA* during interval A, clustering into Group 9; however, during interval B, *ccoN* was quickly up-regulated during the aerobic phase and slightly down-regulated anaerobically, placing it into response Group 4 (CHB2, Group 4 in Fig. 4-7).

The up-regulation of *coxAB* anaerobically combined with aerobic up-regulation of *ccoN* may be a coordinated strategy to counter the repetitive changes between aerobic and anoxic environments and rapidly utilize aerobic machinery as soon as oxygen becomes available as a competitive advantage in microaerobic environments [56]. Remarkably, both terminal oxidase operons (*coxBA* and *ccoN*) also contained a positive FNR binding site upstream, which is consistent with the report on the *cbb3*-type oxidase, indicating that FNR may in fact be a “master-switch” between anaerobic and aerobic metabolism in CHB.

Overall, the 15 genes related to aerobic and anaerobic respiration, discussed previously, were clustered into 6 expression groups (Group 1, 3, 4, 5, 6 and 9). Eight genes (53%) contained a predicted FNR motif upstream of the transcription start site with an FNR site score greater than 5 and these eight genes were clustered with other genes of similar expression pattern in five out of the six expression groups discussed (Group 1, 3, 4, 5 and 9). Group 1 expression was characterized by overall steady down-regulation; Group 3 by little change in regulation between minutes 3-7 but with maximum up-regulation occurring between minute 7-13 followed by steady anaerobic down-regulation; Group 4 had maximum up-regulation occurring at minute 7 (fast) and little change in regulation anaerobically; Group 5 by steady aerobic up-regulation followed by anaerobic down-regulation; Group 6 by sharp aerobic down-regulation and no significant change in regulation anaerobically; and Group 9 by overall steady up-regulation.

Of the 1,169 DEGs in CHB, 586 were significantly expressed or repressed (> median absolute LFC) with 26 of these related to dissimilatory and assimilatory nitrogen

metabolism. Moreover, 21 out of 26 genes involved in nitrogen metabolism had LFC greater than the median and a positive score for FNR direct regulon prediction, suggesting a potential influence by the FNR transcription factor in the CHB nitrogen metabolism.

The AOB metagenome

The assembled AOB metagenome was 2.76 Mbp with GC content of 50.5% and based on presence of single copy genes was determined as 99% complete and 0.7% redundant (Table 4-2). AOB had an identical GC to *Nitrosomonas europaea* (AL954747). AOB contained 2,489 predicted coding sequences with 647 not identified and labeled as hypothetical proteins (of which nearly half were assigned a COG category). AOB also had 156 CDS that were predicted N-cycling gene homologs (~6%). Overall, 2,112 of AOB CDS were assigned to COG categories (85%) and 1,267 were assigned to SEED subsystems (51%).

The phylogenetic classification of AOB as a *Nitrosomonas europaea*-like organism correlated well with the ANI and AAI calculations. The ANI and fraction alignment for 16 *Nitrosomonas* genomes (Fig. 4-9) indicated that AOB is a representative of the *Nitrosomonas europaea* species (ANI 97.9%; fraction aligned 83.8%). This was also confirmed with the high AAI of 97.3% between AOB and *Nitrosomonas europaea* (Fig. S4-3). We also compared AOB with *Nitrosomonas europaea* OLB-2 genome obtained from the Olburgen bioreactor [17], and we observe even greater similarities (ANI 98.8%; fraction aligned 92.3%; AAI 98.8%) (Fig. S4-4). None of the other ANI values with the

AOB and *Nitrosomonas europaea* OLB-2 genome were greater than 75%, further supporting their classification as *Nitrosomonas europaea* strains.

Our pangenomic analysis of the 17 neighboring *Nitrosomonas* (*N.*) genomes from the phylogenetic tree (Fig. 4-3; Fig. 4-9) with a total of 51,137 gene calls resulted in 14,303 gene clusters. AOB genes were grouped into 2,426 clusters, which was very similar to *N. europaea* with 2,456 gene clusters. We grouped all gene clusters into 8 bins based on their occurrence across genomes: (1) Overall core gene clusters (n=965), (2) gene clusters shared exclusively in all genomes within the *N. europaea* lineage (n=59), (3) shared gene clusters between AOB and *N. europaea* (n=322), (4) gene clusters shared between AOB and *N. eutropha* (n= 9), (5) gene clusters shared between AOB and *N. europaea*, and *N. eutropha* (n= 144), (6) unique gene clusters to AOB (n= 237) (6) other singletons (n=7,867), (7) other gene clusters that do not fit any of these classes (n=4,700), (Fig. 7).

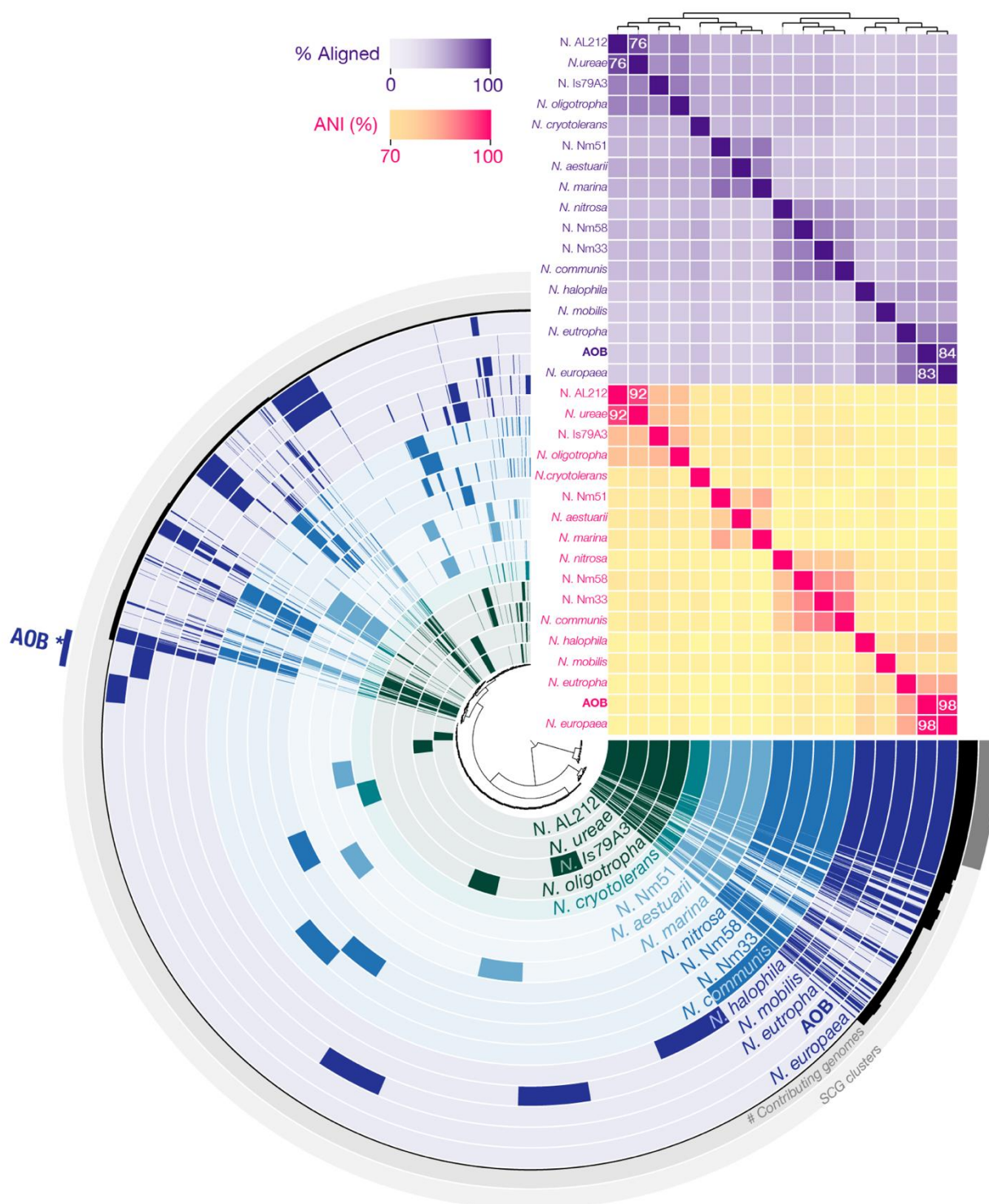


Figure 4-9. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the AOB metagenome (this study) and 16 published *Nitrosomonas* genomes from the phylogenetic tree. All genomes were greater than 95% complete with less than 4% redundancy. Each ring represents a

bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 14,303 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: 'Core genes' (gene clusters shared between all genomes); four 'Shared' bins (gene clusters shared between AOB and *Nitrosomonas europaea* [AL954747], AOB and *Nitrosomonas eutropha* [NC_008344], shared between all three of the former genomes, and shared between all *N. europaea* lineage genomes indicated by the dark blue rings); 'Unique' (gene clusters present in only AOB, indicated on the outside of all rings); 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices.

As expected, the AOB metagenome encoded for a cytochrome aa3-type terminal oxidase gene cluster (*coxCBA*), along with genes encoding for ammonia and hydroxylamine oxidation, *amoCAB* and *hao*, respectively (Fig. 4-5). The median absolute LFC of all differentially expressed genes in AOB was 2.2, meaning that median gene expression was either increased or decreased by a multiplicative factor of $2^{2.2}$ or ~ 4.6 . The gene clusters encoding the terminal oxidase, *amo*, and *hao* were within the *Nitrosomonas* core genome. The gene encoding *coxA* (gene caller ID of 1347) was differentially expressed with significant change (absolute LFC of 4.8) and were clustered into Group 2 (AOB 3, Group 4 in Figure 4-7; Supplementary Table S4-6). The Group 4 expression pattern was characterized by a fast aerobic up-regulation with maximum expression at minute 7 and subsequent maintenance of expression across the aerobic period. Consistent with the findings in CHB, we also detected an FNR binding site upstream of the *coxA* gene in AOB.

The single gene encoding for *hao* (gene caller ID 1698) was differentially expressed, but changes in expression (absolute LFC ~ 2) were below the median absolute LFC. Regardless, the expression pattern of this gene varied slightly from interval A to interval B. In interval A, the *hao* expression was up-regulated steadily until minute 13, and then down-regulated steadily until the end of the interval; this expression profile clustered with the Group 3 response type (AOB 4, Group 3 in Figure 4-7). During interval B, the *hao* expression increased rapidly during the aerobic period without significant changes in expression during the anoxic period, which clustered closely within the same group as the AOB *coxA* gene (AOB 3, Group 4 in Figure 4-7). The continued expression of the cytochrome aa3-type oxidase and hydroxylamine oxidoreductase may support the

maintenance of aerobic respiratory machinery, to be immediately utilized at the next exposure to aerobic conditions in order to increase fitness in the transient microaerobic conditions [56]. In this case, we did not confidently detect an FNR binding site upstream of *hao*, suggesting that the regulation of this gene may be through a different mechanism or the FNR site score of 5 may need to be adjusted for higher specificity to each organism.

AOB contained one operon with the complete *amoCAB* gene set (gene caller IDs of 801, 802, and 803, respectively), with a mean absolute LFC of ~3.1. The expression profiles of *amoC* and *amoB* genes clustered into Group 9 (labeled as AOB 2, Group 9 in Fig. 4-8) together with the *coxAB* and *nosZ* genes from CHB. Interestingly, the expression of genes within the AOB 2 subgroup did not significantly change during the aerobic period between 3 and 7 minutes; however, gene expression was up-regulated during the anoxic conditions. This observation may be explained by a similar hypothesis to the *coxAB* genes in CHB where the microorganism may prepare in advance for a rapid competitive advantage when oxygen becomes available, possibly anticipating the next exposure to oxygen since the bioreactor frequently switches between aerobic and anoxic environments [59]. A similar observation has been reported in other *Nitrosomonas europaea*, where these types of responses are thought to be a strategy to sustain machinery in order to increase fitness during periods of substrate starvation and stress [21]. We detected an FNR binding site upstream of *amoCAB*, *coxA*, and most *nosZ* homologs. The potential FNR activation of *amo* and *coxA* indicates that FNR may also be the mechanism that switches the AOB metabolism between anaerobic and aerobic;

where aerobic metabolism is the preferred option due to the increased energy production in the form of ATP.

The AOB metagenome in this study includes genes encoding both the copper-containing (*nirK*) and haem-containing (*nirS*) nitrite reductases which reduce nitrite to nitric oxide, a cytochrome c-dependent nitric oxide reductase (*cnorB*) which reduces nitric oxide to nitrous oxide, and a nitrous oxide reductase (*nosZ*) [7] (Fig. 4-5). All genes in this denitrification pathway were differentially expressed (FDR < 0.05) with varying dynamics (i.e. absolute LFC). Similar to the observation in CHB, we found several genes with positive hits to *nirK*. Two *nirK* homologs (gene caller ID 1091 and 2077) had identical structure (100% identity over 100% coverage) to a *nirK* protein in the NCycDB database [32], both were located in gene clusters within the *Nitrosomonas* core genome, and both were differentially expressed with significant dynamic change in expression (absolute LFC of 3.8 and 3.0 respectively). Only one of these two *nirK* homologs (gene caller ID 2077) was observed with an FNR motif and the expression profile was similar to those of *amoC* and *amoB* in Group 9 (AOB 2, Group 9 in Fig. 4-8). The other *nirK*-1091 homolog clustered in a different expression response type in Group 1 (AOB 1, Group 1 in Fig. 4-7). This Group 1 expression profile is characterized by little to no change in aerobic expression with a steady decline in expression during the anoxic period. We also found 10 genes similar to *nirS*, with one *nirS* homolog having identical structure to a *nirS* protein in NCycDB (gene caller ID = 2474; 100% identity and 100% coverage) and within a gene cluster in the *Nitrosomonas* core genome; however, the dynamic expression of this gene was below the median absolute LFC (~0.8) and the expression profile did not significantly cluster with any other N-cycling genes. The most

dynamic potential *nirS* homolog (gene caller ID = 960) was also found in the *Nitrosomonas* core genome and had an absolute LFC of approximately 4.2. The expression of *nirS*-960 rapidly increase during the aerobic phase and was followed by minimal change in anoxic expression, clustering closely with both *coxA* and *hao* genes from AOB in response Group 4 (AOB 3, Group 4 in Figure 4-7). The up-regulation of *nirS* indicates that it may have functional relevance in the AOB anaerobic metabolism.

Additionally, AOB encoded for reduction of nitric oxide to nitrous oxide through *norBC*. Neither *norB* nor *norC* were within the *Nitrosomonas* core genome. From the pangenome, the gene cluster was binned in the 'Other' category. These genes had moderate change in expression with absolute LFC around the median absolute LFC of approximately 2. Once again, the gene expression of both the catalytic subunit *norB* (gene caller ID = 2082) and *norC* were steadily up-regulated during the anoxic phase and clustered with Group 9 (AOB 2, Group 9 in Fig. 4-8) along with *amoC*, *amoB*, and *nirK*-2077. The *norBC* gene cluster differed from others in this response group in that we did not confidently detect an FNR binding site upstream, either because it is absent or because the cutoff value used is too stringent for the AOB MAG.

Unlike previous reports, the *Nitrosomonas europaea*- like AOB in this bioreactor not only encoded for, but actively transcribed the genes required for a denitrification pathway complete from nitrite to nitrogen gas. While transcriptional activity is only a proxy for true function, the identification of a nitrous oxide reductase in a *Nitrosomonas europaea* strain living in cyclic microaerobic and anoxic conditions is significant. For over three decades, scientists and engineers have debated the factors leading to nitrous oxide production from N-cycling bacteria such as the ammonia

oxidizing *Nitrosomonas* [60-63]. Some have suggested that these emissions were because the nitrous oxide reductase (NOS) enzyme was inhibited due to a greater sensitivity to oxygen, pH, and sulfide versus nitrogen-oxide reductase enzymes [64]. Several have proposed that oxygen stress promote the use of alternative electron acceptors (nitrite) leading to partial nitrifier denitrification to nitrous oxide [60, 61, 63, 64]; however, Stieglmeier et al. found no effect of decreasing O₂ levels on N₂O production [65]. A specific report on the whole *Nitrosomonas europaea* genome suggests nitrous oxide production may be due to the lack of nitrous oxide reducing machinery altogether, and no full ORFs were identified with strong similarity to nitrous oxide reductase [66]. Metagenomic analysis of other anammox bioreactors also suggest that the specific *Nitrosomonas europaea* strains lack the *nosZ* gene and, thus, cannot complete denitrification by reducing nitrous oxide to nitrogen gas (such as *Nitrosomonas* OLB2 in the Olburgen system) [17]. Clearly, the *nosZ* debate is ongoing, and will likely be so for some time.

Nevertheless, we identified *nosZ* homologs that were not only in the AOB MAG but also part of the *Nitrosomonas* core genome, suggesting that these other *Nitrosomonas* genomes also encode for these particular *nosZ* proteins. We focus on two homologs: one *nosZ* homolog with identical structure to a *nosZ* protein in NCycDB (gene caller ID = 931; 100% identity and 100% coverage) and one *nosZ* homolog with high dynamic change (gene caller ID=587). The *nosZ*-931 gene was differentially expressed (FDR < 0.05), with dynamic change (absolute LFC ~ 2.6) above the median. This *nosZ*-931 was rapidly up-regulated during the aerobic phase of interval A, and expression maintained anaerobically, which again clusters the expression profile into response Group 2 (AOB

3, Group 4 of Figure 4-7) along with the AOB *coxA*, *hao*, and *nirS* genes. During interval B, *nosZ*-931 shared similar expression patterns with response Group 9, which was characterized by steady up-regulation over the entire interval. Like *coxA*, we also detected an FNR binding site upstream of *nosZ*-931. The most dynamic, differentially expressed, potential *nosZ* homolog had an absolute LFC nearly three times greater than the median (gene caller ID = 587; absolute LFC ~6.1). In both intervals, *nosZ*-587 was rapidly up-regulated like *nosZ*-931 and was similarly clustered into response Group 2 (AOB 3, Group 4 of Figure 4-7). However, we did not confidently detect an FNR binding site upstream of *nosZ*-587. The significant differential expression and up-regulation activity of the AOB *nosZ* homologs suggest that this MAG may indeed have the capability to reduce nitrous oxide to nitrogen gas as part of its anaerobic metabolism, and the similarity in expression profiles with other N-cycling genes is evident; however, due to the inconsistency in FNR motif site scores, it is unclear if FNR regulates the expression of *nosZ* in the AOB MAG.

The NOB metagenome

The presence of nitrite oxidizing bacteria in anammox systems is generally considered unfavorable to optimal inorganic nitrogen removal in these systems, since these organisms directly compete with the ammonia oxidizing bacteria for oxygen and with anammox for nitrite. Thus, when nitrite oxidizers persist, they produce nitrate and reduce the overall nitrogen removal efficiency. However, some reports argue that NOB may instead play a part in a beneficial ‘nitrite-loop’ along with heterotrophic denitrifiers in these PNA systems, where the nitrate produced by NOB is partially denitrified back to nitrite and available for use by anammox [17, 18, 67].

The assembled NOB metagenome was 3.82 Mbp with GC content of 60.4% and based on presence of single copy genes was determined as 97% complete and 1.4% redundant (Table 2). NOB had a similar GC content to *Nitrospira defluvii* (NC014355) and contained 3,643 predicted coding sequences with 1,041 unidentified and labeled as hypothetical proteins. However, a third of these hypothetical proteins were recognized by at least one COG category. NOB also had 212 genes related to N-cycling (~6%). Overall, 2,799 NOB genes were assigned to at least one COG category (77%) and 1,281 to SEED subsystems (35%).

Although NOB appeared to cluster near *Nitrospira defluvii* in the phylogenetic tree (Fig. 4-3), we were surprised to find the calculated ANI and fraction alignment for the 7 *Nitrospira* genomes (Fig. 4-10) did not clarify the classification of the NOB genome. The ANI and fraction aligned for NOB and *Nitrospira defluvii* was merely 75.8% and 63.2%, respectively. None of the other ANI values with the NOB were above the 76%, indicating that NOB may be a distinct NOB species. The AAI comparison between NOB and *Nitrospira defluvii* gave a similar result with just 78.2% identity between the two genomes (Fig. S4-5). Since our other key genomes appeared to closely relate with those obtained from the other anammox bioreactors, we decided to compare NOB with the single NOB genome obtained from the Olburgen bioreactor, *Nitrospira* OLB-3 [17]. Remarkably, NOB and *Nitrospira* OLB-3 were nearly identical (ANI 99.6%; fraction aligned 93.3%; AAI 99.5%) (Fig. S4-6), supporting the classification as a *Nitrospira* species.

Our pangenomic analysis of the 8 neighboring *Nitrospira* (N.) genomes from the phylogenetic tree (Fig. 4-3; Fig. 4-10) with a total of 31,020 gene calls resulted in

12,155 gene clusters. NOB genes were grouped into 3564 clusters. We grouped all gene clusters into 5 bins based on their occurrence across genomes: (1) Overall core gene clusters (n=1402), (2) gene clusters shared exclusively between NOB and *N. defluvii* (n=720), (3) unique gene clusters to NOB (n= 510) (6) other singletons (n=7,224), (7) other gene clusters that do not fit any of these classes (n=2,809), (Fig. 8).

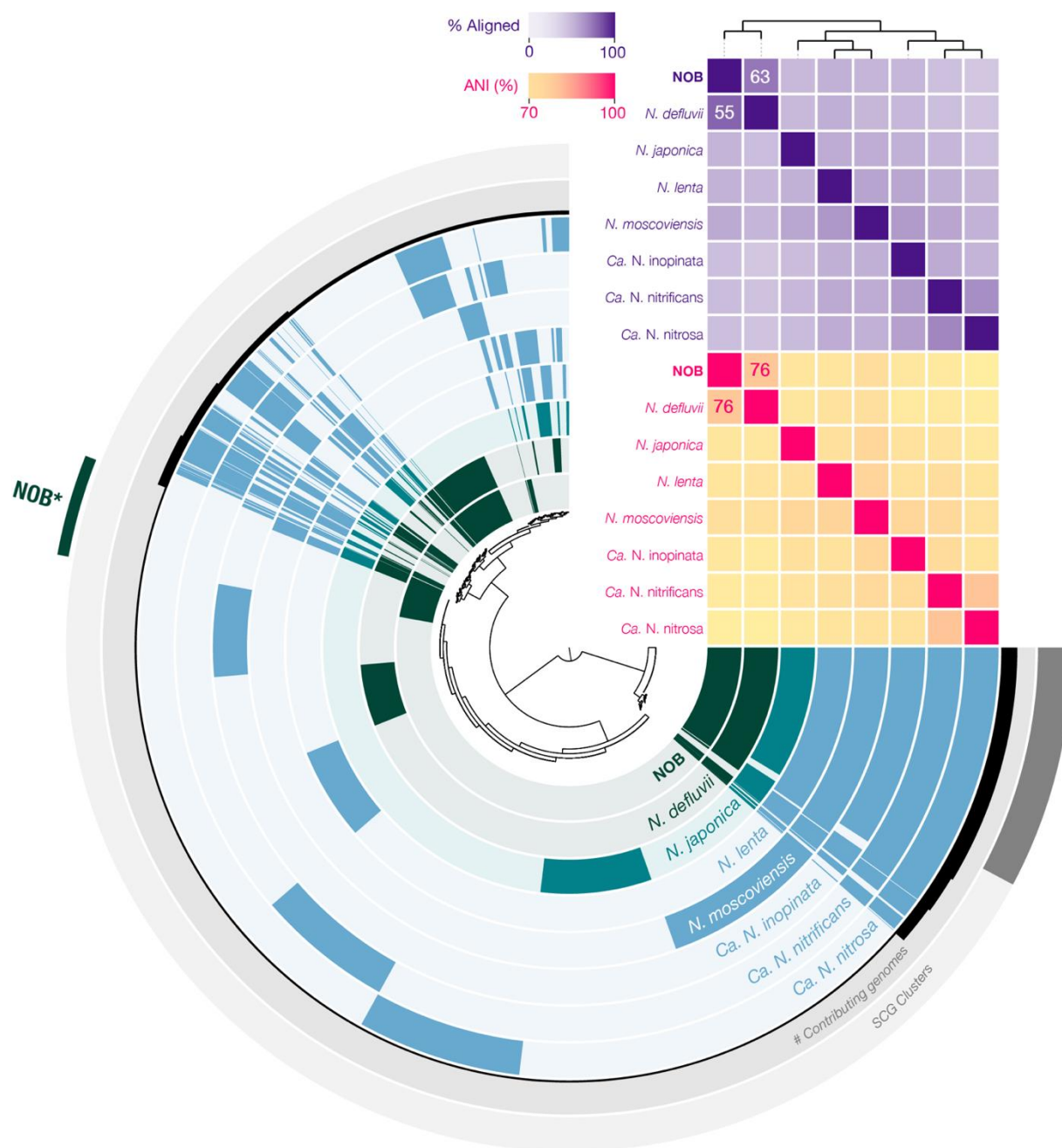


Figure 4-10. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the NOB metagenome (this study) and 7 published *Nitrospira* genomes in the phylogenetic tree. All genomes were greater than 97% complete with less than 5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 12,155 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: ‘Core genes’ (gene clusters shared between all genomes), ‘Shared (gene clusters present in NOB and

closely related genome *Nitrospira defluvii* [NC074355]), 'Unique' (gene clusters present in only NOB, indicated on the outside of all rings), 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices.

The NOB metagenome encoded for a terminal cytochrome bd-I ubiquinol oxidase (*cydAB*), along with the nitrite oxidoreductase (*nxrB*), which oxidizes nitrite to nitrate (Fig. 4-5). The median absolute LFC of all differentially expressed genes in NOB was 3.1. The gene clusters encoding the terminal oxidase and *nxr* were within the *Nitrospira* core genome. The gene encoding *cydA* (gene caller ID of 3047) was differentially expressed but change in expression was below the median (absolute LFC of 2.6), which does not mean that NOB does not utilize this terminal oxidase but instead may not change the expression of *cyd* over the changing conditions. The *cydA* did not contain a predicted FNR site upstream of the transcription start site. However, the *cydA* gene expression profile still clustered significantly with other N-cycling genes in Group 4 (NOB 1, Group 4 in Figure 4-7; Supplementary Table S4-7). The *nxrB* gene (gene caller ID 3050) did not colocalize with an FNR binding site either, and dynamic change (absolute LFC of 1.4) was also below the median; however, the expression profile clustered significantly with that of *cydA* into response Group 4 (NOB 1, Group 4 in Fig. 4-7). The expression pattern of Group 4 is characterized by fast upregulation during the aerobic period and slow down regulation during the anaerobic period, following the same pattern as CHB subgroup 2, which also contained genes encoding for aerobic metabolism (*coxBA*, *ccoN*) and anaerobic metabolism (*nirS*, *nosZ*) - all of which colocalized with FNR motifs in CHB. Therefore, it is possible that *cydA* and *nxrB* are also globally regulated by FNR and the site-score of 5 is too stringent for FNR motif detection in the NOB MAG. Or, an alternative theory could be that the similarities in expression profiles of NOB *cydA* and *nxrB* with the genes from CHB is purely coincidental and the collaborated function of these two enzymes, along with apparent lack of an

FNR binding site, indicates that another regulatory mechanism may control the NOB switch to aerobic metabolism.

The NOB metagenome in this study also included differentially expressed genes for an anaerobic lifestyle through denitrification. NOB encoded for the membrane bound dissimilatory nitrate reductase genes (*narGHJI*) that reduce nitrate to nitrite, and for the both the copper-containing (*nirK*) and haem-containing (*nirS*) nitrite reductases which reduce nitrite to nitric oxide. NOB did encode for the hybrid cluster protein (*hcp*), which is thought to reduce nitric oxide to nitrous oxide, but the genes differential expression was not significantly different from 0 (FDR > 5). We did, however, find a gene encoding for nitrous oxide reductase (*nosZ*), which reduces nitrous oxide to nitrogen gas (Fig. 5). We found two genes encoding for *narG* homologs, both of which were a part of the core *Nitrospira* pangenome. However, the expression profiles of these genes did not cluster significantly with any other N-cycling genes.

Similar to the observation in all other MAGs in this study, we found several genes that were similar to *nirK*. One *nirK* homolog (gene caller ID 1337) had identical structure (100% identity over 100% coverage) to a *nirK* protein in the NCycDB database [32], was differentially expressed but with dynamic change (absolute LFC ~1.58) lower than the median absolute LFC, and was located in a gene cluster within the *Nitrospira* core genome. However, the expression profile of this *nirK* homolog did not cluster significantly with any other N-cycling gene. Another predicted *nirK* homolog (gene caller ID 1531) had the highest dynamic change (LFC ~8), but it was not part of the core genome and the expression profile also did not cluster significantly with any other N-cycling gene. One other predicted *nirK* homolog (gene caller ID = 34) was also highly

dynamic (LFC ~6.6), was part of the core *Nitrospira* pangenome, and clustered significantly with response Group 4 (NOB 1, Group 4 in Fig. 4-7) along with the genes encoding *cycA* and *nxrB*. The strong aerobic up-regulation of this *nirK* may be an indication of its use during aerobic respiration, oxidizing nitric oxide to nitrite as has been suggested previously [7]. The up-regulation may also be a response to nitrite accumulation during the aerobic period, for subsequent nitrite reduction anaerobically; or, perhaps, for both purposes.

In the NOB MAG we also found several genes similar to *nirS*, which were differentially expressed: one potential *nirS* homolog contained the highest percent identity to a *nirS* protein in the NCycDB database (*nirS*-111) and was part of the core *Nitrospira* pangenome and one with the highest dynamic change (*nirS*-234) that binned into the ‘other’ pangenome category. The expression of the *nirS*-111 homolog did not change significantly (~0.95) and was nearly a third less than the median absolute LFC of all genes in NOB. However, the expression profile did cluster significantly into Group 7 (NOB 7, Group 7 in Fig. 4-8). Group 7 is characterized by slight aerobic down-regulation between minutes 3 and 7, followed by up-regulation that reaches a maximum level at minute 13 and subsequent down-regulation during the rest of the anaerobic period. The expression of the *nirS*-234 homolog was highly dynamic (absolute LFC ~ 5.9), and its expression profile clustered significantly into Group 4 (NOB 1, Group 4 in Fig. 4-7). Both of these *nirS* homologs had an FNR binding site predicted upstream. The clustering of the highly dynamic *nirS*-234 into the oxygen-responsive Group 4, with *nirK*, *cycA*, and *nxrB* may suggest that the transcription of all four genes are co-regulated, perhaps by an FNR-type regulator.

Finally, we analyze the differentially expressed *nosZ* homologs identified in the NOB MAG. One gene encoding *nosZ* had the highest percent identity to a *nosZ* protein in NCycDB (*nosZ*-1289), was part of the *Nitrospira* core pangenome, and contained a predicted FNR motif upstream of its transcription start site. Another gene encoding *nosZ* had the highest dynamic change (*nosZ*-2609) with a very high absolute LFC of ~ 9.3 , but did not colocalize with a predicted FNR motif. While the expression of *nosZ*-1289 was below (LFC ~ 1.4) the median absolute LFC of NOB, its expression profile clustered significantly with other N-cycling genes in Group 10 (NOB 4, Group 10 in Fig. 4-8). The Group 10 response type is characterized by aerobic down-regulation between minutes 3 and 7, followed by immediate up-regulation across the entire anaerobic period. Interestingly, even though we did not confidently predict an FNR binding site upstream of *nosZ*-2609, it follows the Group 4 oxygen-response type, just like *nirS*-234, *nirK*-34, *cycA*, and *nxrB* (NOB 1, Group 4 in Fig. 4-7). In either case, the up-regulation of *nosZ* indicates that it may have functional relevance in the NOB metabolism, and possibly beneficial for reducing the emissions of the dangerous greenhouse gas from bioreactors like these.

The AMX metagenome

The assembled AMX metagenome was 3.16 Mbp with GC content of 45% and based on presence of single copy genes was determined as 96% complete and 0.7% redundant (Table 4-2). AMX had an identical GC to the anammox *Candidatus* (*Ca.*) *Brocadia fulgida* (LAQJ01). AMX contained 2,788 predicted coding sequences with 776 not identified and labeled as hypothetical proteins (of which nearly half were assigned a COG category). AMX also had 176 CDS that were predicted N-cycling gene homologs ($\sim 6\%$). Overall,

2,149 of AMX CDS were assigned to COG categories (77%) and 990 were assigned to SEED subsystems (36%).

The phylogenetic classification of AMX as a *Candidatus* (*Ca.*) *Brocadia fulgida*-like organism (Fig. 4-3) was confirmed with the ANI and AAI calculations. The ANI and fraction alignment for 5 anammox genomes (Fig. 4-11) supported the AMX classification as a *Ca. Brocadia fulgida* strain (ANI 99.5%; fraction aligned 91.0%). This was also confirmed with the high AAI of 98.7% between AMX and *Ca. Brocadia fulgida* (Fig. S4-3). Interestingly, although NOB, CHB, and AOB were highly similar to the organisms found in the Olburgen bioreactor, the AMX genome did not closely relate with the anammox genome from the Olburgen system (*Ca. Brocadia sinica* OLB-1) [17]. AMX and *Ca. Brocadia sinica* OLB-1 had only 62.2% ANI and 76.2% AAI, suggesting they are entirely different species. Instead, AMX was also virtually identical to the *Ca. Brocadia* UTAMX2 genome recovered from the laboratory-scale anammox bioreactor in Utah (AAI of 98%; Fig. S9) [18]. We also evaluated the AAI of *Ca. Brocadia* UTAMX2 to *Ca. Brocadia fulgida* (Fig. S4-10), which was approximately 97%. None of the other ANI values with the AMX and other anammox genomes were greater than 75%, supporting the classification as a strain of *Ca. Brocadia fulgida*.

Our pangenomic analysis of the 6 neighboring *Planctomycetes* genomes from the phylogenetic tree (Fig. 4-3; Fig. 4-11) with a total of 18,318 gene calls resulted in 7,706 gene clusters. AMX genes were grouped into 2,667 clusters. We grouped all gene clusters into 6 bins based on their occurrence across genomes: (1) Overall core gene clusters, present in all genomes (n=1151), (2) gene clusters shared exclusively in AMX and *Ca. Brocadia fulgida* (n=377), (3) shared gene clusters between AOB and *Ca.*

Brocadia sinica (n=12), (4) unique gene clusters to AMX (n=173) (6) other singletons (n=4,135), (7) other gene clusters that do not fit any of these classes (n=1,858), (Fig. 4-11).

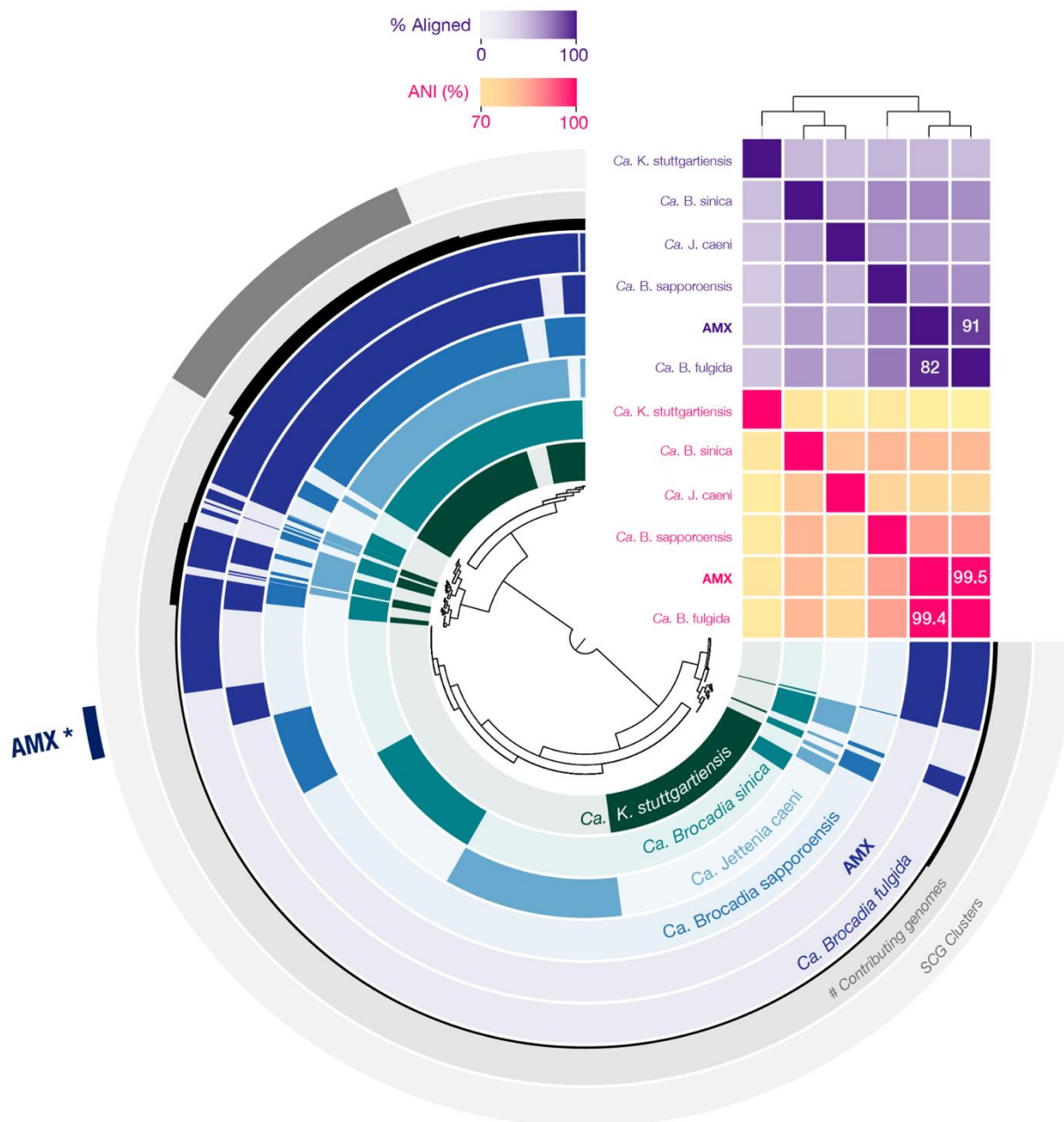


Figure 4-11. Combined pangenome, average nucleotide identity (ANI) and fractioned aligned comparison between the AMX metagenome (this study) and 5 published anammox genomes from the phylogenetic tree. All genomes were greater than 91% complete with less than 5% redundancy. Each ring represents a bacterial genome with shaded and blank regions in each circle representing presence and absence of a particular gene cluster (GC) in that genome, respectively. A total of 7,706 GCs were observed across all genomes. Gene clusters were manually binned into categories based on presence/absence including: ‘Core genes’ (gene clusters shared between all genomes), ‘Shared’ (gene clusters present in AMX and closely related genome *Ca. Brocadia fulgida* [LAQJ01]), ‘Unique’ (gene clusters present in only AMX,

indicated on the outside of all rings), 'Singletons' (other GCs present in only one genome) and 'Other' (everything else). Single copy gene (SCG) clusters are shaded on the outermost ring. ANI values above 90% and alignment above 75% are indicated in their respective matrices. Hierarchical clustering of genomes based on presence/absence of genomic content is indicated by the dendrogram above the ANI and percent aligned matrices.

As expected, the AMX metagenome encoded for and differentially expressed genes required for the anammox metabolism: a single hydrazine synthase (*hzsCBA*) which oxidizes ammonia to hydrazine and requiring the reduction of nitrite by *nirK* or *nirS*, and a single hydrazine dehydrogenase (*hdh*) which oxidizes hydrazine to nitrogen gas (Fig. 4-5). The median absolute LFC of all differentially expressed genes in AOB was 0.9. The gene clusters with the hydrazine synthase and hydrazine dehydrogenase were within the *Planctomycetes* anammox core genome. Only *hzs* expression changed above the median absolute LFC, denoting that *hdh* expression may be sustained throughout each interval. The gene encoding *hzsC* (gene caller ID of 2270) was more dynamic (absolute LFC of 1.26) and clustered into two different response types for interval A and interval B. During the first interval, *hzsC* clustered into Group 6 (AMX 7, Group 6 in Figure 4-8; Supplementary Table S4-8). The Group 6 response type is characterized by an initial aerobic down-regulation followed by little change in expression anaerobically. During interval B, *hzsC* expression clustered into the Group 4 response type (AMX 8, Group 4 in Fig. 4-8), which follows a much different expression pattern by a fast aerobic up-regulation and sustained anaerobic expression. The difference in *hzsC* expression during the two sampling intervals is fascinating and may indicate that anammox were not utilizing ammonia during the aerobic period prior to the end of the bioreactor feeding period (Fig. 4-1) and that heterotrophic nitrite denitrification was the primary pathway for N-removal between hours 3-3.5 of bioreactor operation. Alternatively, this may indicate the presence of some inhibitory substance in the struvite harvesting reject water (feed to the bioreactor), that is degraded by other organisms prior to hours 4.5-5 of bioreactor operation. Evidence that organic carbon is inhibitory to anammox

bacteria has been frequently reported [68-72]. The feed to the MAD MA-AX bioreactor contained variable concentrations of Biochemical Oxygen Demand (BOD) from week to week, ranging from as low as 150 mg O₂/L to as high as 2,300 mg O₂/L. When the RNA-seq samples were collected, the BOD in the feed was nearly 740 mg O₂/L; therefore, it is very likely that the anammox population was initially inhibited until the organic carbon was degraded by heterotrophic denitrifiers (during the first four hours of bioreactor operation), when organic carbon reached a sufficiently low concentration for anammox activity to continue. CHB is probably the biggest contributor to this process, as it is the most abundant and active organism in the bioreactor. Neither *hzcC* nor *hdh* were predicted to have an FNR binding site upstream, but the clustering of *hzcC* expression into Group 4 suggests co-regulation with genes within other microorganisms in the bioreactor that were identified to have an FNR binding motif, thereby supporting the idea that FNR may very well be a global regulator in this microbiome.

Unlike the anammox microorganism from the Olburgen system, the AMX metagenome here also encoded the copper-containing and haem-containing nitrite reductases, *nirK* and *nirS*, respectively (Fig. 4-5). The *nirK* homolog with the greatest dynamic change (LFC ~3; gene caller ID= 2534), was also part of the *Planctomycetes* anammox core pangenome and its expression pattern from interval B clustered with the Group 8 response type (AMX 3, Group 8 in Fig. 4-8). The Group 8 expression patterns are characterized by an initial down-regulation until minute 13, then subsequent up-regulation until the end of the anaerobic period. The expression pattern of *nirK*-2534 during interval A did not significantly cluster with any of the gene expression profiles. The *nirK* homolog with the greatest percent identity (gene caller ID = 165) to a *nirK*

protein had expression dynamics that were greater than the median absolute LFC (absolute LFC ~1.2) and was expressed differently in intervals A and B. During interval A, *nirK*-165 followed the Group 8 expression pattern, like *nirK*-2534, but during interval B *nirK*-165 expression was steadily down-regulated (AMX 6, Group 1 in Fig. 4-7). Two *nirS* homologs were also identified in AMX (Fig. 5), and both were within the *Planctomycetes* anammox core pangenome. The expression of one *nirS* (gene caller ID = 381) was dynamic (absolute LFC ~2.2) and clustered with both *nirK*-2534 and *nirK*-165 in the Group 8 response type (initial down-regulation followed by up-regulation) during interval A (AMX 4, Group 8 in Fig. 4-8), but expression during interval B did not cluster significantly with any of the N-cycling genes. The expression of the other *nirS* (gene caller ID = 1789; absolute LFC ~1.1) clustered with the Group 4 response type during interval B only (AMX 8, Group 4 in Fig. 4-7), which was quickly up-regulated during the aerobic period and expression relatively unchanging anaerobically. Perhaps, this transcriptomic response indicates that *nirS*-1789 encodes for the enzyme that provides nitric oxide for hydrazine synthase activity in AMX. We were not able to confidently detect an FNR binding site in either *nirK* or *nirS* in AMX, however, the expression profile of *nirS*-1789 is highly similar to *nirS* genes in both CHB and NOB that did contain a predicted FNR motif upstream of the transcription start site. This combined with the oxygen-responsive co-expression of genes from different microorganisms, supports the hypothesis that FNR globally regulates *nirS* in this microbiome.

AMX encoded for a hybrid cluster protein *hcp* which was differentially expressed, dynamic (absolute LFC ~1.6; gene caller ID = 460), and part of the *Planctomycetes* anammox core pangenome. The expression profile of *hcp* from interval B clustered with

the Group 7 response type (AMX 1, Group 7 of Fig. 4-7), and indicated that AMX may reduce nitric oxide— perhaps for detoxification purposes. Further, AMX also encoded the nitrous oxide reductase *nosZ*, for reducing nitrous oxide. Five potential *nosZ* homologs were identified in the AMX metagenome and were part of the *Planctomycetes* anammox core pangenome. However, only two of these were dynamic above the median LFC. The expression profile of both *nosZ* homologs clustered in the Group 8 response type (Fig. 4-8). One *nosZ* homolog (gene caller ID = 1612) had an absolute LFC of approximately 2.4, and its expression from both interval A and B was clustered significantly into subgroup AMX 3 and AMX 4, both within response Group 8, suggesting that AMX participated in nitrous oxide reducing activity both before and after the bioreactor was filled. Group 8 is characterized by an initial down-regulation between minutes 3-13 followed by subsequent anaerobic up-regulation between minutes 13-27. The up-regulation of *nosZ* in AMX may be activated by the presence of nitrous oxide, by a distinct regulatory mechanism from FNR.

Remarkably, for the obligate anaerobe, the AMX metagenome contained an aerobic *cbb3*-type terminal oxidase (*ccoN*), which was differentially expressed, part of the *Planctomycetes* anammox core pangenome, and had an FNR binding site predicted upstream of the transcription start site (absolute LFC ~0.7; gene caller ID = 321), (Fig. 4-5). The expression of *ccoN* in AMX followed the Group 9 response type (AMX 2, Group 9 in Fig. 4-8), and was primarily up-regulated during the anaerobic period. This finding is consistent with other reports [16] indicating that anammox in fact contain an ‘unprecedented collection of cytochrome c proteins’ which allow for a metabolic versatility that is not yet understood. It remains unclear if the up-regulation of *cbb3*-

type terminal oxidases in anammox contributes to energy metabolism or is simply used for oxygen detoxification. Either way, this serves as a strategy for AMX to counter the consistently varying environments in the bioreactor. The rapid aerobic transcriptional up-regulation of the *cbb3*-type terminal oxidase in AMX is consistent with the *cbb3*-type terminal oxidase found in CHB, the cytochrome *aa3*-type terminal oxidase in AOB, and the cytochrome *bd-I* ubiquinol oxidase found in NOB— with the majority of these genes containing an FNR motif. This finding along with the oxygen-dependent changes in transcript abundance provides additional evidence for global FNR regulation of the switch between aerobic and anaerobic metabolism in four distinct microorganisms.

Methods

Sequencing batch reactor description and operation

A laboratory-scale bioreactor was originally inoculated with biomass obtained from the full-scale Hampton Roads Sewerage District York River Treatment Plant (Seaford, VA), which uses a deammonification (DEMON®) process to treat reject water from the side stream solids dewatering facility [73]. The DEMON® process is considered one of the most popular sequencing batch reactor (SBR) configurations for achieving partial nitrification/anammox (PN/A) and is unique for its selective enrichment of granular anaerobic ammonium oxidizing bacteria (anammox) biomass via hydro-cyclones (Lackner et al., 2014). DEMON® processes are also operated with intermittent aeration, and aeration periods are controlled based on pH (Wett, 2006). In contrast, the lab-scale bioreactor in this study was designed and operated to simulate a CANON-type process ('Completely Autotrophic Nitrogen removal Over Nitrite')[74] for the sidestream

treatment of the reject water from a full-scale struvite recovery process at the Nine Springs Wastewater Treatment Plant (Madison, WI). This bioreactor did not require pH-based DO control, pH control, or solids retention time (SRT) control.

Effluent (20-L) from the full-scale struvite recovery process at the Nine Springs Wastewater Treatment Plant was collected weekly, stored at 4° C, and was fed to the laboratory-scale bioreactor. Chemical composition of the feed varied from week-to-week depending on full-scale plant operations and performance of upstream processes, with average ammonium concentration of 200 ± 60 mg $\text{NH}_4^+\text{-N/L}$ and variable organic carbon concentrations (soluble chemical oxygen demand, sCOD) ranging primarily between 60 and 400 mg $\text{O}_2\text{/L}$, with a few weeks exceeding 1,000 mg $\text{O}_2\text{/L}$ (Table 1).

The bioreactor was operated with three 8-h cycles per day. Each cycle consisted of a 4-h fill period, where approximately ~1-L of struvite-recovery effluent was fed to the bioreactor, followed by a 3.5-h post-fill period. Subsequently, the biomass settled for 3-min before ~1-L of supernatant was decanted for 12-min. Then, the reactor remained idle for 15-min before beginning the next 8-h cycle. While we did not implement SRT control by intentionally wasting, the rapid settling time likely washed-out small bacterial floc and retained larger aggregates. DO was controlled using an interval-aeration strategy during the 7.5-h fill/post-fill period (repeating 10-min on, 20-min off) (Fig. 4-1; Fig. 4-2). In the 10-min microaerobic period, a LABVIEW (National Instruments, Austin, TX) program was used to automatically turn on/off air-delivery in order to maintain oxygen concentrations below 0.2 mg $\text{O}_2\text{/L}$. The low oxygen concentrations and high temperature are reported to allow ammonia oxidizing bacteria (AOB) to outcompete nitrite oxidizing bacteria (NOB) for available oxygen, allowing

nitrite to accumulate for anammox to use during the anoxic period [75]. After 10-min, air delivery was completely deactivated for the 20-min anoxic period. The pH was uncontrolled but remained at an average of 7.5 ± 0.3 due to the equilibrium between the alkalinity of the feed and acid produced during nitrification (Table 1). A magnetic stir bar and stir plate was used to mix the reactor between 250-300 rpm in the 7.5-h fill/post-fill period. To maintain temperature between 30-35° C, heated water was pumped through flexible tubing wrapped around the exterior of the bioreactor using a Masterflex peristaltic pump (Cole Parmer, Vernon Hills, IL, USA). This reactor sustained PNA activity for 3 years and is regarded to as the MADison Micro-Aerobic/AnoXic (MAD MA-AX) system.

16S ribosomal RNA gene tag sequencing

To classify and determine the relative abundance of bacteria in the reactor, we amplified and sequenced of the V3-V4 region of the 16S rRNA gene were performed. Forty-seven biomass samples were collected from the MAD MA-AX bioreactor sludge between days 60 and 455 of reactor operation. Biomass was saved via centrifugation with a relative centrifugal force of 10,000g for 10 min at to form a biomass pellet, with supernatant discarded, and the pellet frozen at -80 °C. Biomass samples were stored at -80°C until DNA extraction. DNA for 16S ribosomal RNA gene tag sequencing was extracted using DNeasy® PowerSoil® DNA Isolation Kit (Qiagen, Hilden, Germany) following the manufacturer's directions. DNA was quantified with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) and the purity ratio, or ratio of absorbance at 260 nm and 280 nm, was determined with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). DNA samples were stored at -20°C until further

processing. The V3 and V4 regions of the 16S rRNA gene were amplified using the primer set S-D- Bact-0341-b-S-17/S-D-Bact-1061-a-A-17 as described by Klindworth et al. [76]. Amplicons were sequenced on an Illumina MiSeq sequencer (Illumina, San Diego, CA) using pair-end 250 base pair kits at the University of Wisconsin-Madison Biotechnology Center.

The resulting data files containing paired-end reads were quality trimmed and filtered with Sickle Paired End ('sickle pe') using the Sanger fastq file quality type and default minimum length and quality thresholds [77]. The quality trimmed paired-end reads were merged with Fast Length Adjustment of Short Reads (FLASH) version 1.2.11 using default parameters [78]. Then, the merged reads were aligned, filtered, and binned into operational taxonomic units (OTU) with 97% identity using the standard QIIME protocol [79]. USEARCH version 6.1 was used for OTU picking and chimera detection [80]. The representative sequences from each OTU were taxonomically classified using the MIDas-DK database [81].

Metagenomic DNA and RNA sequencing

We collected four 6-ml aliquots of biomass at day 77, day 231, day 350, and day 454 of reactor operations, each of which were centrifuged, decanted, and stored at -80°C until metagenomic DNA extraction. Twenty-four samples were also collected at 454 days, each of which were immediately centrifuged, decanted, flash-frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

For metagenomic DNA extraction, we followed a phenol-chloroform, bead beating protocol for DNA extraction. First, a DNA extraction buffer (pH 8.0) was added to each

sample (500 μ l of 1.5 M sodium chloride, 100 mM Tris, 100 mM ethylenediamine [EDTA], 75 mM sodium phosphate, 1% cetyltrimethylammonium bromide [CTAB]). Next, 30 μ l of lysozyme (100 mg/mL; Thermo Fisher Scientific, MA, USA) was added to each sample and incubated (30 min at 37 °C). Then, 20% sodium dodecyl sulfate [SDS] (10 μ l; BioRad, Hercules, CA) and proteinase K (60 μ l; New England Biolabs, MA, USA) was added and incubated (30 min at 50 °C), followed by another addition of 20% SDS (65 μ l) incubation (30 min at 65 °C). Subsequently, we added a 24:24:1 solution of phenol, chloroform, and isoamyl alcohol (500 μ l) to each sample, subjected them to bead beating (2 min), centrifugation (5,000 relative centrifugal force (rcf); 3 min; room temperature (RT)) and the entire supernatant transferred to a 1.5-ml centrifuge tube. Centrifugation was repeated (12,000 rcf; 10 min; RT) to separate the aqueous and organic phases, the aqueous layer transferred to a new centrifuge tube, and an equal volume of chloroform added to remove residual phenol. We repeated centrifugation (12,000 rcf; 10 min; RT) to re-separate the phases, transferred the aqueous upper layer into a new tube, applied isopropanol (500 μ l) to precipitate nucleic acids, and incubated the samples (minimum of 20 min; 4 °C). Samples were centrifuged (12,000 rcf; 20 min; 4 °C), decanted, washed with 70% ethanol (500 μ l), and air dried (~20 min). The pellets were resuspended in Tris-EDTA buffer (100 μ l) and 10 mg/ml RNase (2 μ l), followed by incubation (15 min; 37 °C). The solution of phenol, chloroform, and isoamyl alcohol (24:24:1) was added again to each sample (100 μ l), centrifuged (12,000 rcf; 10 min; RT), and the aqueous phase transferred to a new tube. This time, chloroform was added to each sample (100 μ l), with centrifugation and aqueous phase separation repeated. Finally, we added 3 M sodium acetate (10 μ l) and 95% ethanol (250 μ l), followed by another incubation step

(24 h at -20 °C), centrifugation (12,000 rcf; 30 min; 4 °C), washing with 70% ethanol, and air-drying (~20 min). Lastly, each sample was resuspended in Tris-EDTA buffer (50 µl). Finally, the DNA was quantified, purified, and assessed with a Qubit 4 Fluorometer (Thermo Fisher Scientific, MA, USA), a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA), and gel electrophoresis.

Prior to RNA extraction, all work surfaces and pipettes were treated with RNaseZap® (Thermo Fisher Scientific, MA, USA). Then, a lysis solution (400 µl) was added to each sample (20 mM sodium acetate, 1 mM EDTA, and 0.5% SDS prepared in water treated with diethylpyrocarbonate [Invitrogen, CA, USA]) followed by incubation (2 min; 65 °C). Next, TRIzol (1 ml; Invitrogen, CA, USA) was added and samples subjected to bead beating using lysing matrix A (MP Biomedicals, CA, USA) for 2 min. After this step, successive phase separation steps performed with mixtures of phenol, chloroform, isoamyl alcohol, and chloroform were used to separate nucleic acids from additional cell material, as described previously. RNA was further purified with an RNEasy minikit (Qiagen, Hilden, Germany) and on-column DNase 1 (Qiagen, Hilden, Germany) treatment. After resuspension of the RNA, quantity, purity, and quality were assessed with a Qubit 4 fluorometer (Thermo Fisher Scientific, MA, USA), a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA), and gel electrophoresis. RNA samples were submitted to the University of Wisconsin Gene Expression Center for quality control with a Bioanalyzer (Agilent, CA, USA), and rRNA reduction was performed with a RiboZero- Bacteria rRNA removal kit (Illumina, CA, USA) with a 1-µg RNA input. Strand-specific cDNA libraries were prepared with a TruSeq RNA library preparation kit (Illumina, CA, USA).

Both DNA and RNA were sequenced with an Illumina HiSeq 2500 platform (Illumina, CA, USA). For DNA, an average insertion size of 550 bp was used and 2×250 -bp reads were generated. For RNA, 1×100 -bp reads were generated. Raw DNA and cDNA read data can be found on the National Center for Biotechnology Information (NCBI) website (see below).

Metagenomic data preparation, quality filtering, read recruitment, taxonomy, and phylogeny

DNA samples from the four sampling days were sequenced twice, resulting in 8 DNA sequencing read files. Low quality DNA sequencing reads were removed using Sickle with a minimum quality score of 20 and a minimum sequence length of 100 [77]. Reads from all eight files were then co-assembled using SPAdes v. 3.3.0 in metaSPAdes mode with kmer values of 21, 33, 55, 77, 99, and 127 [82].

Metagenomic data was binned using Anvi'o v. 5.5.0, following a modified version of the workflow described by Eren et al. [19] (<http://merenlab.org/2016/06/22/anvio-tutorial-v2>). Briefly, after simplifying the assembly scaffolds headers with 'anvi-script-reformat-fasta', the reads that passed quality control criteria were mapped back to the assembly using the bbwrap option of BBMap v. 38.22 [83], with mapping files converted to BAM file format, sorted, and indexed using SAMtools v. 1.9 [84]. Afterward, a contigs database was generated for each of the four samples included in the co-assembly using 'anvi-gen-contigs-database', which also calls open reading frames using Prodigal v. 2.6.3 [26]. Single-copy bacterial and archaeal genes were identified using HMMER v. 3.2.1 [39]. NCBI Clusters of Orthologous Groups (COGs) of proteins were profiled and

included in the contigs databases [35]. A profile for each sample was constructed with contigs >2.5 kbp using ‘anvi-profile’, which hierarchically clusters contigs based on their tetra-nucleotide frequency profiles. The individual profiles were then merged with ‘anvi-merge’, where scaffolds were automatically clustered using a Euclidean distance and Ward linkage algorithm in addition to the CONCOCT algorithm (which uses contig coverage and composition) [85]. Finally, we manually binned the clusters with the Anvi’o interactive interface by running ‘anvi-interactive’. The quality, completeness, and redundancy of each bin was evaluated with ‘anvi-summarize’ and CheckM v. 1.0.3 [86]. We then individually evaluated each initial bin and attempted to manually refine them using ‘anvi-refine’ and added an ‘R’ to the end of the bin name if it required additional refinement. Ultimately, we were able to recover 77 MAGs with >90% completion and <5% redundancy. We performed a final mapping of all MAGs to calculate the mean coverage and detection and determined relative abundance by normalizing the total number of mapped reads for each MAG by the genome size. The metagenomic assembly summary and statistics are summarized in Supplementary Table S2. Supplementary Table S4 reports the genomic features (including completion and redundancy values) of the characterized MAGs.

Taxonomy for each MAG was tentatively evaluated using Kaiju v. 1.7.0 [87] with the NCBI BLAST non-redundant protein database [29] and CheckM [86]. We used PhyloSift v. 1.0.1 with default parameters to infer associations between MAGs and 84 publicly available genomes of related organisms, based on 37 universal single-copy marker genes [20]. Then, we used RAxML v8.2.12 with the PROTGAMMAAUTO model and 100

bootstraps to generate a maximum likelihood phylogenetic tree from the PhyloSift concatenated amino acid alignment [88].

Average nucleotide identity (ANI) for the MAGs versus external genomes was also computed with both ‘anvi-compute-ani’ within Anvi’o, which uses PyANI [89], and the JSpecies webtool [23, 90] based on BLAST+ [27]. Pairwise average amino acid identity (AAI) comparisons were performed using both best hits and reciprocal best hits, one-way AAI and two-way AAI, respectively [52].

Genome annotation and homology search

The four genomes with the largest degree of transcriptomic coverage (CHB, Bin 58; AOB, Bin 48; NOB, Bin 46; AMX, Bin 44) were annotated with Prokka v. 1.13.3 [25], which calls HMMER v. 3.2.1 [39] to search sequences against HMM profile databases, blastp v. 2.7.1+ [27, 91], Aragorn v. 1.2.38 [92] to predict tRNAs, Barrnap v. 0.9 [93] to predict rRNAs, and Prodigal v. 2.6.3 [26] for predicting open reading frames (ORFs). We used Prokka with the --metagenome option to improve gene predictions for fragmented genomes and included a custom reference database. The custom database was created by combining annotated protein sequences from closely related complete reference genomes and protein sequences from NCycDB [32], which integrates N cycle gene families from multiple orthology databases. Briefly, records for complete genomes were downloaded from RefSeq [94] using ‘ncbi-genome-download’ [95] and were subsequently converted to a fasta amino acid database using ‘prokka_genbank_to_fasta_db’. We then added amino acid sequences from NCycDB to the database, and used Cd-hit v. 4.8.1 [96] to cluster all amino acid sequences and

reduce redundancy. The resulting database contained 311,641 sequences and was included with the Prokka annotation pipeline [25]. For comparison purposes, the ORFs were also annotated with RAST with the ‘--reannotate_only’ option to preserve original gene calls [37], SEED subsystems [37], and eggNOG-mapper v. 2 [36].

To check the MAGs for additional nitrogen cycling homologs that may not have been annotated previously, we aligned coding sequences from CHB, AOB, NOB, and AMX separately with NCycDB using DIAMOND v. 0.9.24 [33]. We initially created a DIAMOND reference database from the NCycDB amino acid file with the ‘diamond makedb’ command. Then, the alignment was initiated for each MAG using the ‘diamond blastp’ command with the e-value cutoff specified with ‘-e 0.0001’, the number of target sequences to report alignments set to one with ‘-k 1’, and the ‘--more-sensitive’ option. We considered only hits with an e-value < 10E-10, bit score > 50, percent identity > 30%, and coverage > 80% as cutoffs for gene homology, as recommended by Pearson, 2013 [34].

Pangenome analysis / comparative genomics

To directly compare the genomic content of CHB, AOB, NOB, and AMX with closely related genomes, we followed the Anvi’o pangenome workflow by Delmont and Eren [19, 38]. The metapangenome workflow requires all genomes use the same gene calling method, therefore, after annotating with Prokka and including the custom database mentioned previously, we used ‘anvi-gen-contigs-database’ with the ‘--external-gene-calls’ flag to create an individual database for each genome. Each individual database was then annotated against the COG database [35] and profiled for single-copy genes

using HMMER [39]. Then, the DNA, amino acid sequences, and functional annotations for all the genomes to be compared were stored together with ‘anvi-gen-genomes-storage’. After generating the genome storage file, the ‘anvi-pan-genome’ command was run with the identical parameters (--minbit 0.5, --mcl-inflation 10, --use-ncbi-blast) as recommended for closely related genomes by Delmont and Eren [38]. This pangenome workflow calculates similarities of each amino acid sequence in every genome against every other amino acid sequence using blastp [91]. Weak hits are removed using the ‘minbit’ heuristic [97] (we use the default value of 0.5) and gene clusters are identified using the MCL algorithm [98]. Then, hierarchical clustering of gene clusters was created based on their distribution across genomes based on Euclidean distance and Ward clustering [38]. ANI was calculated between the genomes in the genome storage using ‘anvi-compute-ani’. Then, the pangenome was visualized, and gene clusters binned based on whether they were shared or unique using the ‘anvi-display-pan’ command. Results were exported using ‘anvi-summarize’.

Metatranscriptomic data preparation, quality filtering, read recruitment, and analysis

RNA samples were extracted from triplicate biomass samples collected at 8 time points, resulting in 24 RNA sequencing read files. Low quality RNA sequencing reads were removed using Sickle using the single end option ‘se’ [77]. Read quality was verified with FastQC [99]. We used SortMeRNA to remove rRNA sequences using multiple databases for RNA identification [100], and the remaining non-rRNA sequences were mapped to the draft genomes using BMap v. 35.92 with the minimum sequence identity set to 0.95 and ambiguous mappings randomly assigned [83]. The number of RNA reads

mapping to each ORF within each MAG was calculated with htseq-count v. 0.6.1 with ‘-m intersection-strict -s no -a 0 -t CDS -i ID’ parameters [101]. Relative gene expression estimates were calculated for each MAG based on the total number of reads mapped, normalized by the genome size.

We used the R Bioconductor [102] package DESeq2 [40] to perform differential analysis gene-level count data. We built the DESeq dataset directly from htseq-count data using ‘DESeqDataSetFromHTSeqCount’. We initially prefiltered low count genes (≤ 10) as recommended by the DESeq2 Bioconductor workflow [103]. Then, we performed differential expression analysis using the likelihood ratio test (LRT) to test for significant differential expression across all time points using the ‘DESeq’ function [40]. Finally, we treated each sampling interval separately (4 time points each) and extracted comparisons between each time point to the first aerobic time point. Results tables were generated with log₂ fold changes, p values, and adjusted p values using the function ‘results’ [40]. ‘True’ differential expression was defined by a stringent adjusted P value of false discovery rate (FDR) < 0.05 [103].

Gene expression time series data clustering

To group differentially expressed genes into a small number of response types, we applied a nonparametric model-based method (Dirichlet process Gaussian process mixture model [DPGP]) designed for time series gene expression clustering, which incorporates cluster number uncertainty and time series dependencies [48]. This enabled us to explore if shared regulatory mechanisms lead to shared gene expression patterns. We specifically extracted the log₂fold change data over time for annotated

nitrogen cycling genes or predicted homologs that were also differentially expressed in each MAG (FDR < 0.05). The 'DP_GP_cluster.py' command mean centered and scaled log₂ fold change data (standardized log₂ fold change) by subtracting the mean expression from each point and dividing by the standard deviation. DPGP then clustered genes by expression over the time series, generated plots displaying the cluster mean, 2 x standard deviation, and individual gene trajectory as well as created gene-by-gene posterior similarity matrices [48].

FNR sequence motif identification

To identify a potential regulon of the anaerobic transcription factor FNR in the microbiome, we performed in silico analysis to detect sequences with homology to FNR 300 base pairs upstream of the translation start site of all genes in each of the four MAGs with the greatest transcriptomic coverage.

The prediction of FNR sequence motif identification was performed as described by Myers et al. [4]. Briefly, the MAGs were searched for predicted FNR binding sites using the program Patser [104]. Because FNR binding sites are fairly well conserved throughout bacterial species the *E. coli* pair-weight matrix (PWM) was used [105]. A score threshold of 5 was used as a minimum cutoff value for a positive FNR binding site prediction because it was the approximate numerically calculated cutoff score based on the information content of the PWM as determined by Patser.

Data availability

Raw DNA and cDNA read data can be found on the National Center for Biotechnology Information (NCBI) website under BioProject accession no. PRJNA559529.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

NKB, KSM, and DRN developed the research plan, methodology, and project goals. NKB operated the bioreactor, designed and performed laboratory work, as well as *in silico* metagenome and metatranscriptome analysis. KSM contribute to metatranscriptomic analysis and identification of FNR binding sites. NKB drafted the manuscript, tables and figures. NKB, KSM, MS, and DRN contributed to data analysis and manuscript writing.

Funding

This work was partially supported by funding from the National Science Foundation (CBET-1435661, CBET-1803055) and the Madison Metropolitan Sewerage District (Madison, WI).

References

1. Busby, S.J.W., *Transcription activation in bacteria: ancient and modern*. Microbiology, 2019. **165**(4): p. 386-395.
2. Spiro, S. and J. Guest, *FNR and its role in oxygen-regulated gene expression in Escherichia coli* FEMS Microbiology Reviews, 1990. **6**(4).
3. Körner, H., H.J. Sofia, and W.G. Zumft, *Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs*. FEMS Microbiology Reviews, 2003. **27**(5): p. 559-592.
4. Myers, K.S., et al., *Genome-scale analysis of escherichia coli FNR reveals complex features of transcription factor binding*. PLoS Genet, 2013. **9**(6): p. e1003565.
5. Hartig, E. and D. Jahn, *Regulation of the anaerobic metabolism in Bacillus subtilis*. Adv Microb Physiol, 2012. **61**: p. 195-216.
6. Browning, D., et al. *Secrets of bacterial transcription initiation taught by the Escherichia coli FNR protein*. Cambridge; Cambridge University Press; 1999.
7. Kuypers, M.M.M., H.K. Marchant, and B. Kartal, *The microbial nitrogen-cycling network*. Nat Rev Microbiol, 2018. **16**(5): p. 263-276.
8. van de Graaf, A.A., et al., *Anaerobic oxidation of ammonium is a biologically mediated process*. Applied and Environmental Microbiology, 1995. **61**(4): p. 1246-51.
9. Kuenen, J.G., *Anammox bacteria: from discovery to application*. Nat Rev Microbiol, 2008. **6**(4): p. 320-6.
10. Morales, N., et al., *Integration of the Anammox process to the rejection water and main stream lines of WWTPs*. Chemosphere, 2015. **140**: p. 99-105.

11. Wang, S., et al., *Anaerobic ammonium oxidation in traditional municipal wastewater treatment plants with low-strength ammonium loading: Widespread but overlooked*. Water Res, 2015. **84**: p. 66-75.
12. Ma, B., et al., *Biological nitrogen removal from sewage via anammox: Recent advances*. Bioresour Technol, 2016. **200**: p. 981-90.
13. McCarty, P.L., *What is the Best Biological Process for Nitrogen Removal: When and Why?* Environ Sci Technol, 2018. **52**(7): p. 3835-3841.
14. Gao, H., Y.D. Scherson, and G.F. Wells, *Towards energy neutral wastewater treatment: methodology and state of the art*. Environ Sci Process Impacts, 2014. **16**(6): p. 1223-46.
15. He, S.L., et al., *The Treatment Performance and the Bacteria Preservation of Anammox: A Review*. Water Air and Soil Pollution, 2015. **226**(5).
16. Kartal, B., et al., *Anammox--growth physiology, cell biology, and metabolism*. Adv Microb Physiol, 2012. **60**: p. 211-62.
17. Speth, D.R., et al., *Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system*. Nat Commun, 2016. **7**: p. 11172.
18. Lawson, C.E., et al., *Metabolic network analysis reveals microbial community interactions in anammox granules*. Nature Communications, 2017. **8**: p. 15416.
19. Eren, A.M., et al., *Anvi'o: an advanced analysis and visualization platform for 'omics data*. PeerJ, 2015. **3**: p. e1319.
20. Darling, A.E., et al., *PhyloSift: phylogenetic analysis of genomes and metagenomes*. PeerJ, 2014. **2**: p. e243-e243.
21. Yu, R. and K. Chandran, *Strategies of Nitrosomonas europaea 19718 to counter low dissolved oxygen and high nitrite concentrations*. BMC Microbiol, 2010. **10**: p. 70.

22. Lucker, S., et al., *A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria*. Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13479-84.
23. Richter, M. and R. Rossello-Mora, *Shifting the genomic gold standard for the prokaryotic species definition*. Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19126-31.
24. Konstantinidis, K.T. and J.M. Tiedje, *Towards a genome-based taxonomy for prokaryotes*. J Bacteriol, 2005. **187**(18): p. 6258-64.
25. Seemann, T., *Prokka: rapid prokaryotic genome annotation*. 2014. **30**(14): p. 2068-9.
26. Hyatt, D., et al., *Prodigal: prokaryotic gene recognition and translation initiation site identification*. BMC Bioinformatics, 2010. **11**: p. 119.
27. Camacho, C., et al., *BLAST+: architecture and applications*. BMC Bioinformatics, 2009. **10**: p. 421.
28. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase*. Nucleic Acids Res, 2004. **32**(Database issue): p. D115-9.
29. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Res, 2005. **33**(Database issue): p. D501-4.
30. Bateman, A., et al., *The Pfam protein families database*. Nucleic Acids Res, 2004. **32**(Database issue): p. D138-41.
31. Haft, D.H., et al., *TIGRFAMs and Genome Properties in 2013*. Nucleic Acids Res, 2013. **41**(Database issue): p. D387-95.

32. Tu, Q., et al., *NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes*. Bioinformatics, 2019. **35**(6): p. 1040-1048.
33. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND*. Nature Methods, 2015. **12**(1): p. 59-60.
34. Pearson, W.R., *An introduction to sequence similarity ("homology") searching*. Curr Protoc Bioinformatics, 2013. **Chapter 3**: p. Unit3 1.
35. Galperin, M.Y., et al., *Expanded microbial genome coverage and improved protein family annotation in the COG database*. Nucleic Acids Res, 2015. **43**(Database issue): p. D261-9.
36. Huerta-Cepas, J., et al., *Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper*. Mol Biol Evol, 2017. **34**(8): p. 2115-2122.
37. Overbeek, R., et al., *The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)*. Nucleic Acids Res, 2014. **42**(Database issue): p. D206-14.
38. Delmont, T.O. and A.M. Eren, *Linking pangenomes and metagenomes: the Prochlorococcus metapangenome*. PeerJ, 2018. **6**: p. e4320.
39. Eddy, S.R., *Accelerated profile HMM searches*. PLoS computational biology, 2011. **7**(10): p. e1002195.
40. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. Genome Biol, 2014. **15**(12): p. 550.
41. Oshlack, A. and M.J. Wakefield, *Transcript length bias in RNA-seq data confounds systems biology*. Biol Direct, 2009. **4**: p. 14.

42. Dillies, M.A., et al., *A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis*. Brief Bioinform, 2013. **14**(6): p. 671-83.
43. Wagner, G.P., K. Kin, and V.J. Lynch, *Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples*. Theory Biosci, 2012. **131**(4): p. 281-5.
44. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biol, 2010. **11**(10): p. R106.
45. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
46. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
47. Tielen, P., et al., *Anaerobic Regulatory Networks in Bacteria*, in *Bacterial Regulatory Networks*, A.A.M. Filloux, Editor. 2012, Caister Academic Press: London, UK.
48. McDowell, I.C., et al., *Clustering gene expression time series data using an infinite Gaussian process mixture model*. PLoS Comput Biol, 2018. **14**(1): p. e1005896.
49. Lackner, S., et al., *Full-scale partial nitritation/anammox experiences--an application survey*. Water Res, 2014. **55**: p. 292-303.
50. Abma, W.R., et al., *Upgrading of sewage treatment plant by sustainable and cost-effective separate treatment of industrial wastewater*. Water Sci Technol, 2010. **61**(7): p. 1715-22.
51. Thompson, C.C., et al., *Microbial genomic taxonomy*. BMC genomics, 2013. **14**.

52. Rodriguez-R, L.M. and K.T. Konstantinidis, *Bypassing Cultivation To Identify Bacterial Species*. Microbe, 2014.
53. Dutilh, B.E., et al., *Signature genes as a phylogenomic tool*. Mol Biol Evol, 2008. **25**(8): p. 1659-67.
54. B, S., B. P, and H. MA, *Genome phylogeny based on gene content*. Nature Genetics, 1999.
55. Wang, J., et al., *The roles of the hybrid cluster protein, Hcp and its reductase, Hcr, in high affinity nitric oxide reduction that protects anaerobic cultures of Escherichia coli against nitrosative stress*. Mol Microbiol, 2016. **100**(5): p. 877-92.
56. Beach, N.K., *This was revealed to me in a dream once*. . Me, 2019.
57. Gier, J.-W.L.d., et al., *Structural and functional analysis of aa3-type and cbb3-type cytochrome c oxidases of Paracoccus denitrificans reveals significant differences in proton-pump design*. Molecular Microbiology, 1996. **20**(6): p. 1247-1260.
58. Osamura, T., et al., *Specific expression and function of the A-type cytochrome c oxidase under starvation conditions in Pseudomonas aeruginosa*. PloS one, 2017. **12**(5): p. e0177957-e0177957.
59. Beach, N.K., *Suddenly materialized in my mind*. . Me, 2019.
60. Goreau, T.J., et al., *Production of NO₂⁻ and N₂O by Nitrifying Bacteria at Reduced Concentrations of Oxygen*. Appl Environ Microbiol, 1980. **40**(3): p. 526-532.
61. Poth, M. and D.D. Focht, *¹⁵N Kinetic Analysis of N₂O Production by Nitrosomonas europaea: an Examination of Nitrifier Denitrification*. Applied and Environmental Microbiology, 1985. **49**(5): p. 1134.

62. Anderson, I.C., et al., *A comparison of NO and N₂O production by the autotrophic nitrifier Nitrosomonas europaea and the heterotrophic nitrifier Alcaligenes faecalis*. Applied and Environmental Microbiology, 1993. **59**(11): p. 3525.
63. Zhu, X., et al., *Ammonia oxidation pathways and nitrifier denitrification are significant sources of N₂O and NO under low oxygen availability*. Proc Natl Acad Sci U S A, 2013. **110**(16): p. 6328-33.
64. Zumft, W.G. and P.M.H. Kroneck, *Respiratory Transformation of Nitrous Oxide (N₂O) to Dinitrogen by Bacteria and Archaea*. 2006. p. 107-227.
65. Stieglmeier, M., et al., *Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea*. ISME J, 2014. **8**(5): p. 1135-46.
66. Chain, P., et al., *Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea*. Journal of Bacteriology, 2003. **185**(9): p. 2759-2773.
67. Winkler, M.K., et al., *Unravelling the reasons for disproportion in the ratio of AOB and NOB in aerobic granular sludge*. Appl Microbiol Biotechnol, 2012. **94**(6): p. 1657-66.
68. Chen, C., et al., *Evaluation of COD effect on anammox process and microbial communities in the anaerobic baffled reactor (ABR)*. Bioresour Technol, 2016. **216**: p. 571-8.
69. Joss, A., et al., *Combined nitrification-anammox: advances in understanding process stability*. Environmental science & technology, 2011. **45**(22): p. 9735-9742.
70. Jia, L., et al., *Effect of organic carbon on nitrogen conversion and microbial communities in the completely autotrophic nitrogen removal process*. Environmental Technology, 2012. **33**(10): p. 1141-1149.

71. Ni, S.-Q., et al., *Effect of organic matter on the performance of granular anammox process*. Bioresource Technology, 2012. **110**: p. 701-705.
72. Jin, R.-C., et al., *The inhibition of the Anammox process: A review*. Chemical Engineering Journal, 2012. **197**: p. 67-79.
73. Nifong, A., et al. *Performance of a full-scale sidestream demon installation*. in *Proceedings of the WEF Nutrient Removal and Recovery Conference*. 2013. Vancouver, British Columbia, Canada.
74. Sliekers, A.O., et al., *Completely autotrophic nitrogen removal over nitrite in one single reactor*. Water Res, 2002. **36**(10): p. 2475-82.
75. Perez, J., et al., *Outcompeting nitrite-oxidizing bacteria in single-stage nitrogen removal in sewage treatment plants: a model-based study*. Water Res, 2014. **66**: p. 208-218.
76. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies*. Nucleic Acids Res, 2013. **41**(1): p. e1.
77. Joshi, N. and J. Fass, *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]*. 2011.
78. Magoc, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies*. Bioinformatics, 2011. **27**(21): p. 2957-63.
79. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data*. Nat Methods, 2010. **7**(5): p. 335-6.
80. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST*. Bioinformatics, 2010. **26**(19): p. 2460-1.
81. McIlroy, S.J., et al., *MiDAS: the field guide to the microbes of activated sludge*. Database (Oxford), 2015. **2015**: p. bav062.

82. Nurk, S., et al., *metaSPAdes: a new versatile metagenomic assembler*. Genome Research, 2017. **27**(5): p. 824-834.
83. Bushnell, B., *BBMap: a fast, accurate, splice-aware aligner*. 2014.
84. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
85. Alneberg, J., et al., *Binning metagenomic contigs by coverage and composition*. Nature Methods, 2014. **11**(11).
86. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes*. Genome research, 2015. **25**(7): p. 1043-1055.
87. Menzel, P., K.L. Ng, and A. Krogh, *Fast and sensitive taxonomic classification for metagenomics with Kaiju*. Nat Commun, 2016. **7**: p. 11257.
88. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*. Bioinformatics, 2014. **30**(9): p. 1312-3.
89. Pritchard, L., et al., *Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens*. Analytical Methods, 2016. **8**(1): p. 12-24.
90. Richter, M., et al., *JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison*. Bioinformatics, 2016. **32**(6): p. 929-31.
91. Altschul, S.F., et al., *Basic Local Alignment Search Tool*. Journal of Molecular Biology, 1990. **215**(3): p. 403-410.
92. Laslett, D. and B. Canback, *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences*. Nucleic Acids Res, 2004. **32**(1): p. 11-6.

93. Seemann, T., *Barrnap 0.9: Rapid ribosomal RNA prediction*. <https://github.com/tseemann/barrnap>.
94. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation*. *Nucleic Acids Res*, 2016. **44**(D1): p. D733-45.
95. Blin, K., *ncbi-genome-download 0.2.9* <https://github.com/kblin/ncbi-genome-download/> Accessed June 2019.
96. Li, W. and A. Godzik, *Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences*. *Bioinformatics*, 2006. **22**(13): p. 1658-9.
97. Benedict, M.N., et al., *ITEP: an integrated toolkit for exploration of microbial pan-genomes*. *BMC Genomics*, 2014. **15**: p. 8.
98. van Dongen, S. and C. Abreu-Goodger, *Using MCL to Extract Clusters from Networks*, in *Bacterial Molecular Networks: Methods and Protocols*, J. van Helden, A. Toussaint, and D. Thieffry, Editors. 2012, Springer New York: New York, NY. p. 281-295.
99. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010.
100. Kopylova, E., L. Noé, and H. Touzet, *SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data*. *Bioinformatics*, 2012. **28**(24): p. 3211-3217.
101. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data*. *Bioinformatics*, 2015. **31**(2): p. 166-169.
102. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. *Genome Biol*, 2004. **5**(10): p. R80.

103. Love, M.I., S. Anders, and W. Huber. *Analyzing RNA-seq data with DESeq2*. 2019 [cited 2019; Available from: <https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2>].
104. Hertz, G.Z. and G.D. Stormo, *Identifying DNA and protein patterns with statistically significant alignments of multiple sequences*. Bioinformatics, 1999. 15(7-8): p. 563-77.
105. Eckweiler, D., et al., *PRODORIC2: the bacterial gene regulation database in 2018*. Nucleic Acids Res, 2018. 46(D1): p. D320-D326.

Supplementary Tables

Table S4-1. Samples selected for metagenomic sequencing, respective total inorganic nitrogen (TIN) performance, and 16S rRNA sequencing relative abundance from key functional groups.

Common Functional Group:		Anammox	Ammonia Oxidizer	Nitrite Oxidizer	
Day	TIN Removal (%)	<i>Ca. Brocadia</i> (%)	<i>Nitrosomonas</i> (%)	<i>Nitrospira</i> (%)	Other (%)
77	63	1.1	4.5	1.5	93
231	38	0.79	1.6	1.4	96
350	89	2.1	4	0.53	93
454	82	0.91	3.4	0.41	95

Table S4-2. Metagenomic DNA read summary and statistics

Platform	Sample Day	DNA Read File	Raw Reads (R1 + R2)	Post Trimming (Sickle)	Post Error Correction	Mapped to Assembly	% Mapped to Assembly	Reads Assigned to all Anvi'o bins	% Mapped to Anvi'o bins *	Reads Assigned to top 16 bins	% Mapped to top 16 bins *
Hi Seq 2x250	454	2017-05-18_L001	22,228,710	21,813,422	21,733,398	19,699,452	91%	17,937,155	91%	12,113,780	61%
		2017-05-18_L002	21,621,640	21,217,203	21,142,513	19,175,926		17,455,320		11,785,113	
		SUM	43,850,350	43,030,625	42,875,911	38,875,378		35,392,475		23,898,893	
	350	2017-02-03_L001	28,105,414	27,586,992	27,531,158	25,282,849	92%	22,948,459	91%	12,266,318	49%
		2017-02-03_L002	27,140,038	26,635,225	26,583,935	24,421,949		22,155,528		11,842,120	
		SUM	55,245,452	54,222,217	54,115,093	49,704,798		45,103,987		24,108,438	
	230	2016-10-07_L001	27,906,756	27,177,738	27,104,526	24,659,904	91%	21,890,049	89%	9,862,624	40%
		2016-10-07_L002	27,096,116	26,380,649	26,313,514	23,956,192		21,263,212		9,583,152	
		SUM	55,002,872	53,558,387	53,418,040	48,616,096		43,153,261		19,445,776	
	77	2016-05-06_L001	26,478,370	25,909,351	25,836,979	23,360,172	90%	21,137,216	90%	2,965,318	13%
		2016-05-06_L002	25,701,268	25,145,348	25,078,415	22,688,854		20,523,741		2,880,650	
		SUM	52,179,638	51,054,699	50,915,394	46,049,026		41,660,957		5,845,968	
	* % Mapped to bins is based on the number of reads that mapped to the assembly (reads mapped to bins/reads mapped to assembly)										

Table S4-3. Metatranscriptomic cDNA read summary and statistics

Platform	RNA sample	RNA read File	Raw Reads	Post Trimming (Sickle)	Post rRNA Removal	mRNA Mapped to Assembly	% mRNA Mapped to Assembly	Reads Assigned: all Anvi'o bins	% mRNA Mapped to Anvi'o bins *	Reads Assigned: top 16 bins	% mRNA Mapped to 16 focus bins *
HiSeq 1x100	t0733	t0733-1	21,566,227	21,497,432	17,762,013						
		t0733-2	20,411,261	20,341,165	16,763,619	40,369,382	81%	37,142,966	92%	33,185,532	82%
		t0733-3	19,312,652	19,240,705	15,597,957						
		SUM	61,290,140	61,079,302	50,123,589	40,369,382		37,142,966		33,185,532	
	t0737	t0737-1	21,001,100	20,921,575	17,025,012						
		t0737-2	18,673,218	18,605,692	14,631,047	40,745,930	82%	37,448,712	92%	31,857,971	78%
		t0737-3	21,974,163	21,886,835	17,810,261						
		SUM	61,648,481	61,414,102	49,466,320	40,745,930		37,448,712		31,857,971	
	t0743	t0743-1	21,057,201	20,975,827	16,620,008						
		t0743-2	20,214,271	20,137,193	15,824,708	39,738,874	82%	37,218,160	94%	30,844,853	78%
		t0743-3	20,051,398	19,980,294	16,280,516						
		SUM	61,322,870	61,093,314	48,725,232	39,738,874		37,218,160		30,844,853	
	t0757	t0757-1	18,184,249	18,118,284	14,504,457						
		t0757-2	20,122,091	20,067,326	16,450,582	41,083,843	83%	37,874,419	92%	33,104,279	81%
		t0757-3	23,702,578	23,644,554	18,603,162						
		SUM	62,008,918	61,830,164	49,558,201	41,083,843		37,874,419		33,104,279	
	t0903	t0903-1	25,415,759	25,343,114	20,147,885						
		t0903-2	25,413,069	25,343,327	20,269,094	48,950,276	83%	45,138,518	92%	39,317,174	80%
		t0903-3	22,758,236	22,692,418	18,256,229						
		SUM	73,587,064	73,378,859	58,673,208	48,950,276		45,138,518		39,317,174	
	t0907	t0907-1	26,029,431	25,947,005	20,771,156						
		t0907-2	25,790,208	25,721,863	20,896,416	52,828,810	84%	48,976,701	93%	42,572,590	81%
		t0907-3	26,135,302	26,072,740	21,114,290						
		SUM	77,954,941	77,741,608	62,781,862	52,828,810		48,976,701		42,572,590	
	t0913	t0913-1	22,711,481	22,658,786	18,244,395						
		t0913-2	24,681,036	24,619,747	19,388,646	60,437,924	82%	56,295,949	93%	50,250,468	83%
		t0913-3	45,242,916	45,107,244	35,637,548						
		SUM	92,635,433	92,385,777	73,270,589	60,437,924		56,295,949		50,250,468	
	t0927	t0927-1	45,039,580	44,900,049	34,159,409						
		t0927-2	49,133,570	48,985,745	37,935,006	87,423,341	83%	80,721,723	92%	71,698,881	82%
		t0927-3	40,708,595	40,568,670	32,697,868						
		SUM	134,881,745	134,454,464	104,792,283	87,423,341		80,721,723		71,698,881	

* % Mapped to bins is based on the number of reads that mapped to the assembly (reads mapped to bins/reads mapped to assembly)

Table S4-4. Genome statistics of the highly complete/low redundancy MAGs

Bin # *	Taxonomy (family)	Completeness (%)	Redundancy (%)	Genome size (Mb)	# Scaffolds	N50 (scaffolds)	GC (%)	Predicted genes
Bin_58	Ignavibacteriaceae	97	0.0	2.43	19	336,200	38	2,092
Bin_44	Candidatus Brocadiaaceae	96	0.7	3.16	182	27,074	45	2,788
Bin_48	Nitrosomonadaceae	99	0.7	2.76	84	64,454	51	2,489
Bin_46	Nitrospiraceae	97	1.4	3.82	46	130,102	60	3,643
Bin_43	Anaerolineaceae	97	1.4	3.8	262	26,194	57	3,582
Bin_57	Sterolibacteriaceae	100	0.7	3.15	84	82,753	66	3,184
Bin_49	Unk. Bacteroidetes	99	0.7	3.15	38	167,470	36	2,766
Bin_54R	Sterolibacteriaceae	99	1.4	2.88	48	144,614	67	2,873
Bin_38	Unk. Bacteria	98	1.4	7.65	49	347,644	59	5,664
Bin_68R	Ignavibacteriaceae	97	1.4	3.45	19	447,731	42	2,679
Bin_50R	Haliscomenobacteraceae	96	0.0	3.71	83	138,066	48	2,834
Bin_55	Anaerolineaceae	98	1.4	4.06	62	273,379	55	3,681
Bin_56R	Anaerolineaceae	97	2.2	3.79	18	577,897	54	3,515
Bin_32R	Solibacteraceae	94	2.9	5.98	137	70,762	62	5,075
Bin_95R	Unk. Bacteria	99	1.4	9	521	27,066	71	7,129
Bin_99	Unk. Bacteria	96	0.0	3.64	43	213,845	49	3,032
Bin_2	Sinobacteraceae	99	0.0	3.96	164	56,954	66	3,808
Bin_3	Rhodanobacteraceae	100	0.7	3.88	90	70,938	67	3,402
Bin_4R	Unk. Bacteria	96	1.4	2.89	21	302,575	67	2,872
Bin_6R	Unk. Bacteria	94	1.4	3.2	28	145,580	55	2,904
Bin_7R	Unk. Betaproteobacteria	96	1.4	3.68	333	14,790	70	3,782
Bin_9R	Unk. Bacteria	99	0.7	3.62	48	175,022	66	3,287
Bin_12	Unk. Bacteria	96	0.0	2.77	57	203,895	49	2,225
Bin_13	Unk. Bacteria	95	1.4	5.8	464	17,995	53	4,915
Bin_14R	Unk. Bacteria	96	0.7	4.79	93	104,852	65	3,793
Bin_15R	Unk. Bacteria	96	1.4	5.3	60	152,589	67	4,184
Bin_16R	Unk. Bacteria	96	2.2	4.61	311	22,153	73	4,130
Bin_17	Fimbriimonadaceae	96	0.0	3.31	2	3,304,852	59	3,015
Bin_19R	Unk. Bacteria	98	2.9	3.57	27	170,705	55	3,004
Bin_20	Flavobacteriaceae	97	0.7	3.07	83	67,169	33	2,716
Bin_23R	Unk. Bacteria	97	1.4	4.83	36	167,251	60	4,155
Bin_27R	Sterolibacteriaceae	96	2.2	2.96	117	40,656	67	2,968
Bin_29	Haliscomenobacteraceae	99	1.4	7.24	82	194,715	52	5,541
Bin_31	Unk. Bacteria	98	0.7	4.33	40	355,874	62	3,657
Bin_33	Unk. Bacteria	97	0.7	5.23	335	45,529	60	4,818
Bin_34	Unk. Bacteria	97	0.0	4.66	273	28,273	57	4,295
Bin_35	Sinobacteraceae	100	0.7	3.55	45	179,347	68	3,317
Bin_37R	Unk. Bacteria	97	1.4	4.24	41	157,789	55	3,311
Bin_39	Unk. Bacteria	99	1.4	4.87	55	472,994	44	4,308
Bin_40R	Unk. Bacteria	95	1.4	8.85	82	136,258	49	6,196
Bin_41	Nitrosomonadaceae	99	0.0	3.22	97	64,201	49	3,081
Bin_42R	Unk. Bacteria	96	1.4	6.96	314	36,022	58	5,721
Bin_45	Ignavibacteriaceae	99	1.4	3.89	62	146,039	35	3,179
Bin_47R	Anaerolineaceae	94	0.7	3.85	34	167,927	53	3,583
Bin_51	Unk. Bacteria	99	0.0	3.81	51	185,756	66	3,218
Bin_52R	Ignavibacteriaceae	99	1.4	4.06	53	113,126	35	3,521
Bin_53R	Anaerolineaceae	99	0.7	3.53	57	127,505	56	3,242
Bin_62	Unk. Bacteria	95	1.4	2.71	25	1,594,128	62	2,607
Bin_63R	Unk. Bacteria	95	2.2	3.26	4	1,880,758	63	2,686
Bin_64	Ignavibacteriaceae	99	1.4	4.65	67	240,437	44	3,999
Bin_65R	Unk. Bacteria	99	0.7	3.47	30	262,426	34	3,025
Bin_69	Unk. Bacteria	99	0.7	4	96	84,037	68	3,566
Bin_70R	Unk. Bacteria	98	0.7	5.92	64	157,907	63	4,516
Bin_71R	Unk. Bacteria	99	1.4	4.12	108	75,537	63	3,414
Bin_72R	Fimbriimonadaceae	91	0.0	2.87	162	31,835	62	2,727
Bin_75R	Unk. Alphaproteobacteria	100	1.4	3.71	48	124,143	66	3,653
Bin_76R	Unk. Bacteria	98	1.4	5.54	81	105,574	68	4,409
Bin_77	Unk. Bacteria	99	0.7	3.02	107	56,028	64	2,890
Bin_78	Unk. Actinobacteria	99	0.0	4.03	57	112,487	67	3,679
Bin_79R	Unk. Bacteria	96	1.4	5.87	35	291,821	62	4,659
Bin_81R	Unk. Bacteria	93	1.4	7.91	62	252,904	57	6,444
Bin_82R	Unk. Bacteroidia	95	0.0	2.08	37	111,090	49	1,800
Bin_83R	Leptospiraceae	96	2.9	3.1	29	207,588	45	2,920
Bin_85R	Unk. Bacteria	91	3.6	4.86	689	8,561	70	4,383
Bin_87R	Unk. Bacteria	91	1.4	4.49	202	34,753	63	4,078
Bin_89R	Unk. Bacteria	99	0.7	6.19	71	148,422	66	4,785
Bin_90R	Unk. Bacteria	94	1.4	4.36	106	61,818	65	3,714
Bin_91R	Unk. Bacteria	92	1.4	3.78	94	61,889	64	3,181
Bin_96R	Unk. Bacteria	97	1.4	6.65	80	145,076	62	5,476
Bin_97	Fimbriimonadaceae	95	0.7	2.82	48	211,359	60	2,601
Bin_100	Unk. Bacteria	100	0.0	3.28	47	172,435	42	2,878
Bin_121_4R	Unk. Bacteria	96	0.7	2.66	231	20,864	44	2,667
Bin_122_1R	Unk. Bacteria	96	1.4	4.52	127	57,534	68	3,720
Bin_122_2R	Gemmatimonadaceae	95	2.9	3.97	22	386,540	64	3,273
Bin_129_1	Haliscomenobacteraceae	98	0.7	5.75	89	148,391	52	4,496
Bin_130_1	Unk. Bacteria	97	0.0	2.94	67	66,534	39	2,477
Bin_131_1R	Unk. Bacteria	96	2.2	9.07	139	119,210	69	7,387

Table S4-5. Summary of CHB gene expression clustering

MAG ID	Gene homolog	Gene caller id	%ID	%COV	Interval	Pangenome bin	FNR	LFC	Subgroup	Group	Figure
CHB	<i>ccoN</i>	1697			A	SHARED	FNR	1.13	CHB1	9	8
CHB	<i>ccoN</i>	1697			B	SHARED	FNR	1.13	CHB2	4	7
CHB	<i>cox</i>	1882			A	SHARED	FNR	0.82	CHB1	9	8
CHB	<i>cox</i>	1883			A	SHARED	FNR	1.53	CHB1	9	8
CHB	<i>cox</i>	279			A	OTHER		0.8	CHB2	4	7
CHB	<i>nirK</i>	1844	36.2	99.6	B	OTHER		1.35	CHB2	4	7
CHB	<i>nirK</i>	1844	36.2	99.6	A	OTHER		1.35	CHB6	6	8
CHB	<i>nirK</i>	60	31.7	93	A	SHARED		0.94	CHB6	6	8
CHB	<i>nirK</i>	1740	69	98.9	A	OTHER	FNR	0.9	CHB4	3	7
CHB	<i>nirK</i>	1740	69	98.9	B	OTHER	FNR	0.9	CHB4	3	7
CHB	<i>nirS</i>	532	43	81.1	A	SHARED	FNR	1.25	CHB7	1	7
CHB	<i>nirS</i>	532	43	81.1	B	SHARED	FNR	1.25	CHB2	4	7
CHB	<i>nirS</i>	1445	61.6	98.5	A	CORE	FNR	0.69	CHB1	9	8
CHB	<i>nirS</i>	1445	61.6	98.5	B	CORE	FNR	0.69	CHB2	4	7
CHB	<i>narG</i>	741				SHARED		4.43	none	none	none
CHB	<i>narH</i>	742	100	100	A	SHARED		5.33	CHB3	5	7
CHB	<i>narJ</i>	743			A	SHARED	FNR	4.95	CHB3	5	7
CHB	<i>narJ</i>	743			B	SHARED	FNR	4.95	CHB4	3	7
CHB	<i>narH</i>	74	37	96.2	A	OTHER	FNR	3.58	CHB5	8	8
CHB	<i>narH</i>	74	37	96.2	B	OTHER	FNR	3.58	CHB5	8	8
CHB	<i>hcp</i>	80	37.2	99.5	A	SHARED	FNR	2.88	CHB3	5	7
CHB	<i>hcp</i>	80	37.2	99.5	B	SHARED	FNR	2.88	CHB3	5	7
CHB	<i>nosZ</i>	93	100	100	A	OTHER	FNR	0.95	CHB7	1	7
CHB	<i>nosZ</i>	680	35.2	90	A	SHARED		0.59	CHB6	6	8
CHB	<i>nosZ</i>	680	35.2	90	B	SHARED		0.59	CHB1	9	8
CHB	<i>nosZ</i>	1741	39	97	A	OTHER	FNR	0.78	CHB2	4	7
CHB	<i>nosZ</i>	1741	39	97	B	OTHER	FNR	0.78	CHB4	3	7

Table S4-6. Summary of AOB gene expression clustering

MAG ID	Gene Homolog	Gene caller id	%ID	%COV	Interval	Pangenome bin	FNR	LFC	Subgroup	Group	Figure
AOB	<i>cox</i>	1347			A	CORE	FNR	4.82	AOB3	4	7
AOB	<i>cox</i>	1351			-	CORE		3.44	none	none	none
AOB	<i>cox</i>	753			-	OTHER		3.96	none	none	none
AOB	<i>cox</i>	752			-	OTHER		1.36	none	none	none
AOB	<i>amoC</i>	801	99.6	100	B	OTHER	FNR	2.63	AOB2	9	8
AOB	<i>amoA</i>	802			B	CORE	FNR	3.78	AOB2	9	8
AOB	<i>amoB</i>	803	99.8	100	B	CORE	FNR	3.02	AOB2	9	8
AOB	<i>amoC</i>	241			-	OTHER		0.35	none	none	none
AOB	<i>hao</i>	1698	99.8	100	A	CORE		2.04	AOB4	3	7
AOB	<i>hao</i>	1698	99.8	100	B	CORE		2.04	AOB3	4	7
AOB	<i>nirK</i>	1091	100	100		CORE		3.8	AOB1	1	7
AOB	<i>nirK</i>	2077	100	100		CORE	FNR	3	AOB2	9	8
AOB	<i>nirK</i>	791	33.8	94.3	A	OTHER		5.5	AOB3	4	7
AOB	<i>nirK</i>	791	33.8	94.3	B	OTHER		5.5	AOB2	9	8
AOB	<i>nirK</i>	469	42.5	89.9	B	CORE	FNR	4.6	AOB1	1	7
AOB	<i>nirK</i>	1545	30.7	87.1	B	SHARED		3.6	AOB1	1	7
AOB	<i>hcp</i>	226			-	CORE		2.48	none	none	none
AOB	<i>hcp</i>	816	51.1	87.7	A	CORE		0.96	AOB5	10	8
AOB	<i>napA</i>	375			-	CORE		3.37	none	none	none
AOB	<i>nirS</i>	960	61	97.8	A	CORE		4.2	AOB3	2	7
AOB	<i>nirS</i>	961			B	CORE		4.2	AOB3	4	7
AOB	<i>nirS</i>	1986	88	100	A	OTHER		3.97	AOB1	1	7
AOB	<i>nirS</i>	2067	56.7	99.8	B	CORE	FNR	3.48	AOB3	4	7
AOB	<i>nirS</i>	202	78.7	100	B	CORE	FNR	0.86	AOB4	3	7
AOB	<i>nirS</i>	2474			-	CORE		0.81	none	none	none
AOB	<i>norB</i>	2082	100	100	-	OTHER		2	none	none	none
AOB	<i>norC</i>	2083	100	100	A	OTHER		1.89	AOB2	9	8
AOB	<i>norC</i>	2083	100	100	B	OTHER		1.89	AOB3	4	7
AOB	<i>hcp</i>	226				CORE		2.48	none	none	none
AOB	<i>hcp</i>	816	51.1	87.7	A	CORE		0.96	AOB5	10	8
AOB	<i>nosZ</i>	587	32	91.8	A	CORE		6.08	AOB3	4	7
AOB	<i>nosZ</i>	587	32	91.8	B	CORE		6.08	AOB3	4	7
AOB	<i>nosZ</i>	1252	40.4	90.1	A	UNIQUE		3.88	AOB2	9	8
AOB	<i>nosZ</i>	9	38.5	98.5	B	UNIQUE	FNR	3.23	AOB2	9	8
AOB	<i>nosZ</i>	931	100	100	A	CORE	FNR	2.56	AOB3	4	7
AOB	<i>nosZ</i>	931	100	100	B	CORE	FNR	2.56	AOB2	9	8

Table S4-7. Summary of NOB gene expression clustering

MAG ID	Gene Homolog	Gene caller id	%ID	%COV	Interval	Pangenome bin	FNR	LFC	Subgroup	Group	Figure
NOB	<i>cydA</i>	3047	-	-	B	CORE		2.58	NOB1	4	7
NOB	<i>hao</i>	1536	47.5	93.5	A	OTHER	FNR	3.83	NOB1	4	7
NOB	<i>nirK</i>	1337			-	CORE		1.58	none	none	none
NOB	<i>nirK</i>	1531			-	OTHER		7.99	none	none	none
NOB	<i>nirK</i>	34	34.5	89.2	A	CORE		6.55	NOB1	4	7
NOB	<i>nirK</i>	2498	45.3	93.2		SHARED		5.61	NOB5	6	8
NOB	<i>nxr</i>	2650			-	CORE		2.06	none	none	none
NOB	<i>nxr</i>	2556	59	99.7	-	CORE		1.38	none	none	none
NOB	<i>nxr</i>	3050	-	-	A	CORE		1.25	NOB1	4	7
NOB	<i>narG</i>	1591			-	CORE	FNR	2.99	none	none	none
NOB	<i>narG</i>	1547			-	CORE		2.45	none	none	none
NOB	<i>narJ</i>	2901			-	CORE		2.66	none	none	none
NOB	<i>narY</i>	2	33.9	96.4	B	OTHER		3.27	NOB3	8	8
NOB	<i>nirS</i>	233				OTHER	FNR	11.88	none	none	none
NOB	<i>nirS</i>	111	92.8	100	A	CORE	FNR	0.95	NOB7	7	8
NOB	<i>nirS</i>	111	92.8	100	B	CORE	FNR	0.95	NOB7	7	8
NOB	<i>nirS</i>	234	74	98.4	B	OTHER	FNR	5.92	NOB1	4	7
NOB	<i>nirS</i>	2663	31.5	94.1	B	CORE		4.3	NOB3	8	8
NOB	<i>nirS</i>	1763	51.1	97.9	A	CORE		3.53	NOB5	6	8
NOB	<i>nosZ</i>	1281	97.5	100	A	CORE	FNR	1.38	NOB4	10	8
NOB	<i>nosZ</i>	1281	97.5	100	B	CORE	FNR	1.38	NOB4	10	8
NOB	<i>nosZ</i>	862			-	CORE		0.98	none	none	none
NOB	<i>nosZ</i>	2609	36.2	83.4	A	CORE		9.33	NOB1	4	7

Table S4-8. Summary of AMX gene expression clustering

MAG ID	Gene Homolog	Gene caller id	%ID	%COV	Interval	Pangenome bin	FNR	LFC	Subgroup	Group	Figure
AMX	<i>ccoN</i>	321			B	CORE	FNR	0.69	AMX2	9	8
AMX	<i>cydA</i>	2124			B	CORE		1.19	AMX6	1	7
AMX	<i>hao</i>	2192			-	OTHER		1.2	none	none	none
AMX	<i>hao</i>	466	45.4	89.2	A	OTHER	FNR	0.99	AMX6	3	7
AMX	<i>hao</i>	466	45.4	89.2	B	OTHER	FNR	0.99	AMX5	2	7
AMX	<i>hao</i>	24	30.4	86.5	A	CORE		0.58	AMX9	2	7
AMX	<i>nirK</i>	165	89.9	100	A	OTHER		1.21	AMX3	8	8
AMX	<i>nirK</i>	165	89.9	100	B	OTHER		1.21	AMX6	1	7
AMX	<i>nirK</i>	2534	33.8	80.1	B	CORE		3.01	AMX4	8	8
AMX	<i>nirK</i>	2208	30.2	89.2	A	SHARED		0.8	AMX2	9	8
AMX	<i>hzsC</i>	2270	100	100	A	CORE		1.26	AMX7	6	8
AMX	<i>hzsC</i>	2270	100	100	B	CORE		1.26	AMX8	4	7
AMX	<i>hdh</i>	35	40.8	85.3	B	CORE		0.38	AMX3	8	8
AMX	<i>nrfa</i>	1243			-	OTHER		3.29	none	none	none
AMX	<i>nrfaH</i>	1117			A	OTHER	FNR	1.15	AMX4	8	8
AMX	<i>nrfaH</i>	1118			-	OTHER	FNR	0.8	none	none	none
AMX	<i>narG</i>	1306	42.1	99.5	A	CORE		0.82	AMX9	2	7
AMX	<i>narG</i>	144			-	OTHER	FNR	0.75	none	none	none
AMX	<i>narG</i>	332			-	CORE	FNR	0.45	none	none	none
AMX	<i>narH</i>	143	52.9	98.8	B	OTHER	FNR	0.89	AMX6	1	7
AMX	<i>narH</i>	143	52.9	98.8	A	OTHER	FNR	0.89	AMX4	8	8
AMX	<i>narH</i>	329	97	100	B	OTHER	FNR	0.83	AMX4	8	8
AMX	<i>napAB</i>	2716	50.6	98.6	A	CORE	FNR	0.59	AMX3	8	8
AMX	<i>napAB</i>	2717	55.2	99.8	A	OTHER	FNR	0.59	AMX3	8	8
AMX	<i>napAB</i>	2717	55.2	99.8	B	OTHER	FNR	0.59	AMX9	2	7
AMX	<i>napA</i>	2196	37.5	99.8	B	CORE		0.53	AMX1	7	8
AMX	<i>napA</i>	1234			B	CORE		0.25	AMX6	1	7
AMX	<i>nirS</i>	381	61.1	91.7	A	CORE		2.18	AMX4	8	8
AMX	<i>nirS</i>	646	42	99.2	-	OTHER		1.6	none	none	none
AMX	<i>nirS</i>	1789	52.2	99.8	B	CORE		1.07	AMX8	4	7
AMX	<i>hcp</i>	460	63.5	100	B	CORE		1.55	AMX1	7	8
AMX	<i>hcp</i>	2035	36.5	96.7	B	CORE		0.98	AMX4	8	8
AMX	<i>hcp</i>	270	43.5	99.5	A	UNIQUE		0.67	AMX3	8	8
AMX	<i>nosZ</i>	2278	36.3	84.1	A	OTHER	FNR	3.47	AMX7	6	8
AMX	<i>nosZ</i>	2278	36.3	84.1	B	OTHER	FNR	3.47	AMX9	2	7
AMX	<i>nosZ</i>	1612	40.4	100	A	CORE		2.42	AMX4	8	8
AMX	<i>nosZ</i>	1612	40.4	100	B	CORE		2.42	AMX3	8	8
AMX	<i>nosZ</i>	2234	38.7	88.2	B	CORE	FNR	1.48	AMX4	8	8
AMX	<i>nosZ</i>	957	49.1	94.5	A	SHARED	FNR	0.87	AMX6	1	7
AMX	<i>nosZ</i>	961	82.4	100	A	CORE	FNR	0.85	AMX5	3	7
AMX	<i>nosZ</i>	768	72	96.1	A	CORE		0.24	AMX6	1	7

Supplementary Figures

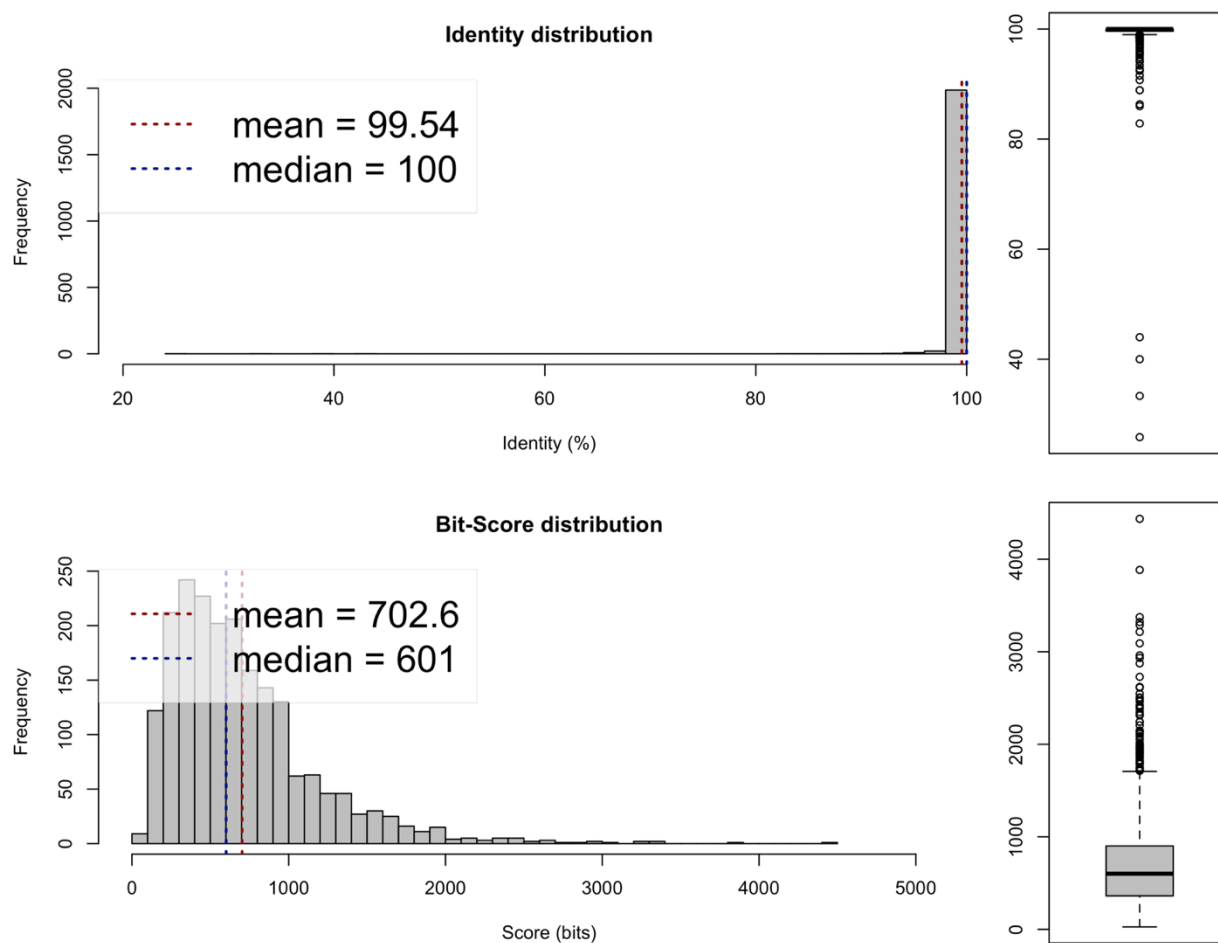


Figure S4-1. Amino acid identity (AAI) and bit score distribution of CHB versus *Chlorobi bacterium OLB4*

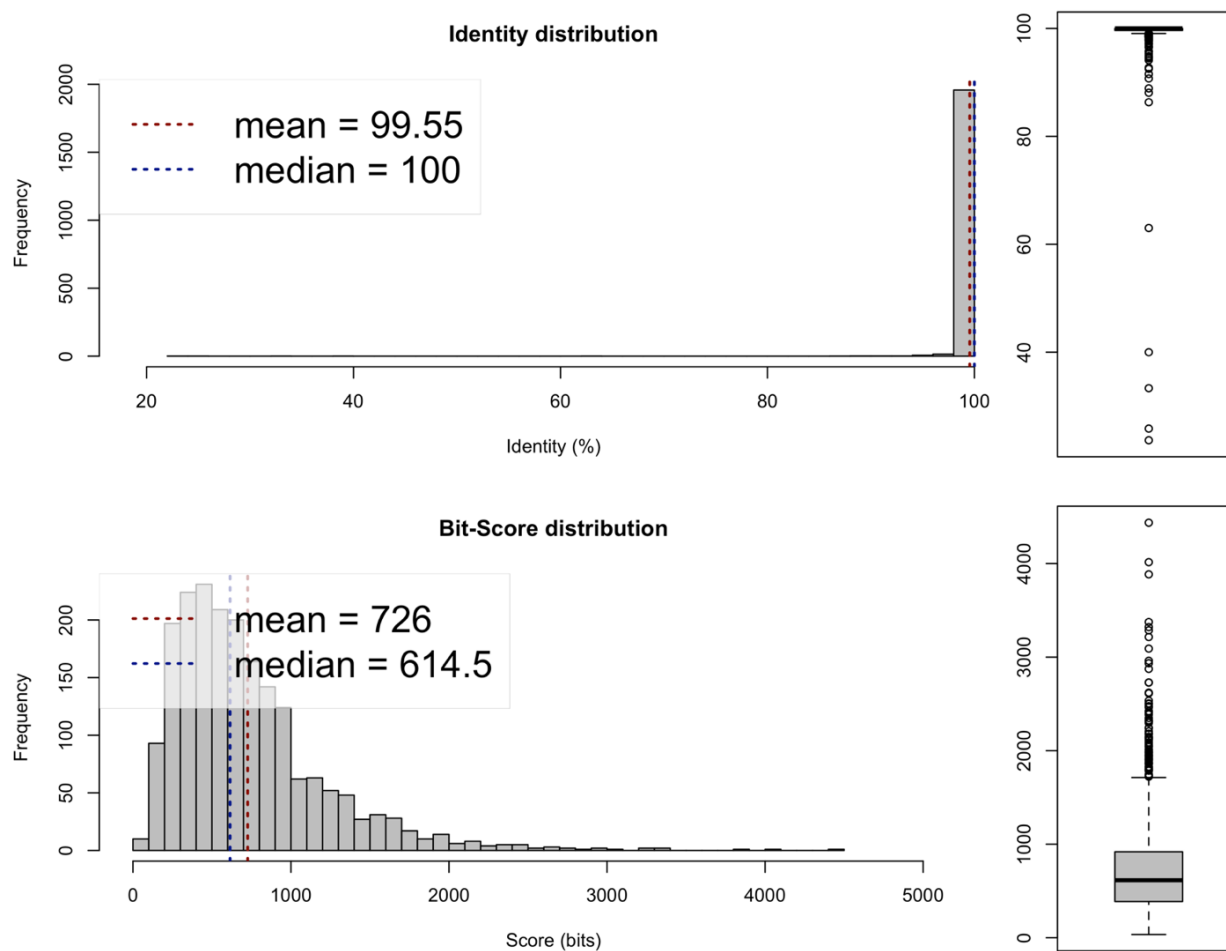


Figure S4-2. Amino acid identity (AAI) and bit score distribution of CHB versus Ignavibacteriales UTCHB1

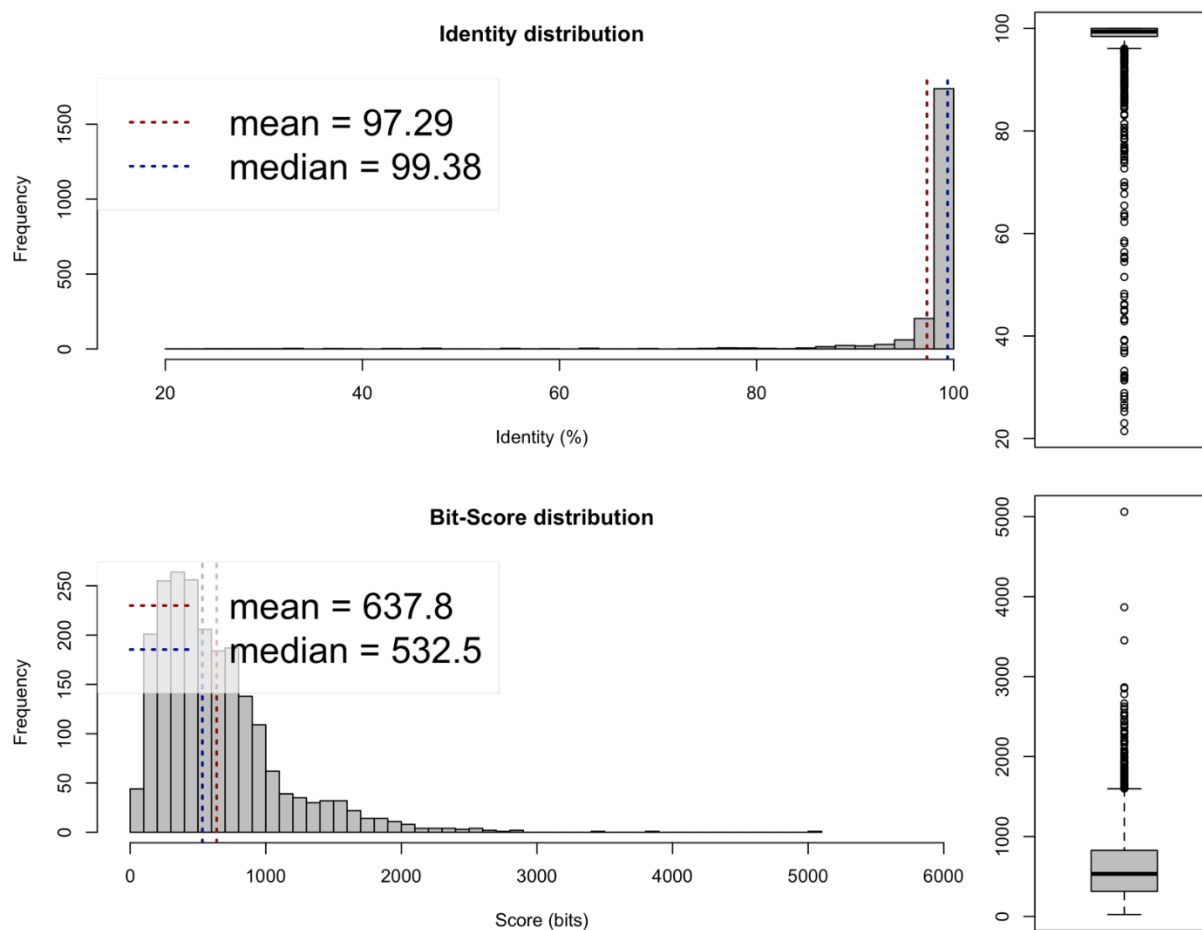


Figure S4-3. Amino acid identity (AAI) and bit score distribution of AOB versus *Nitrosomonas europaea*

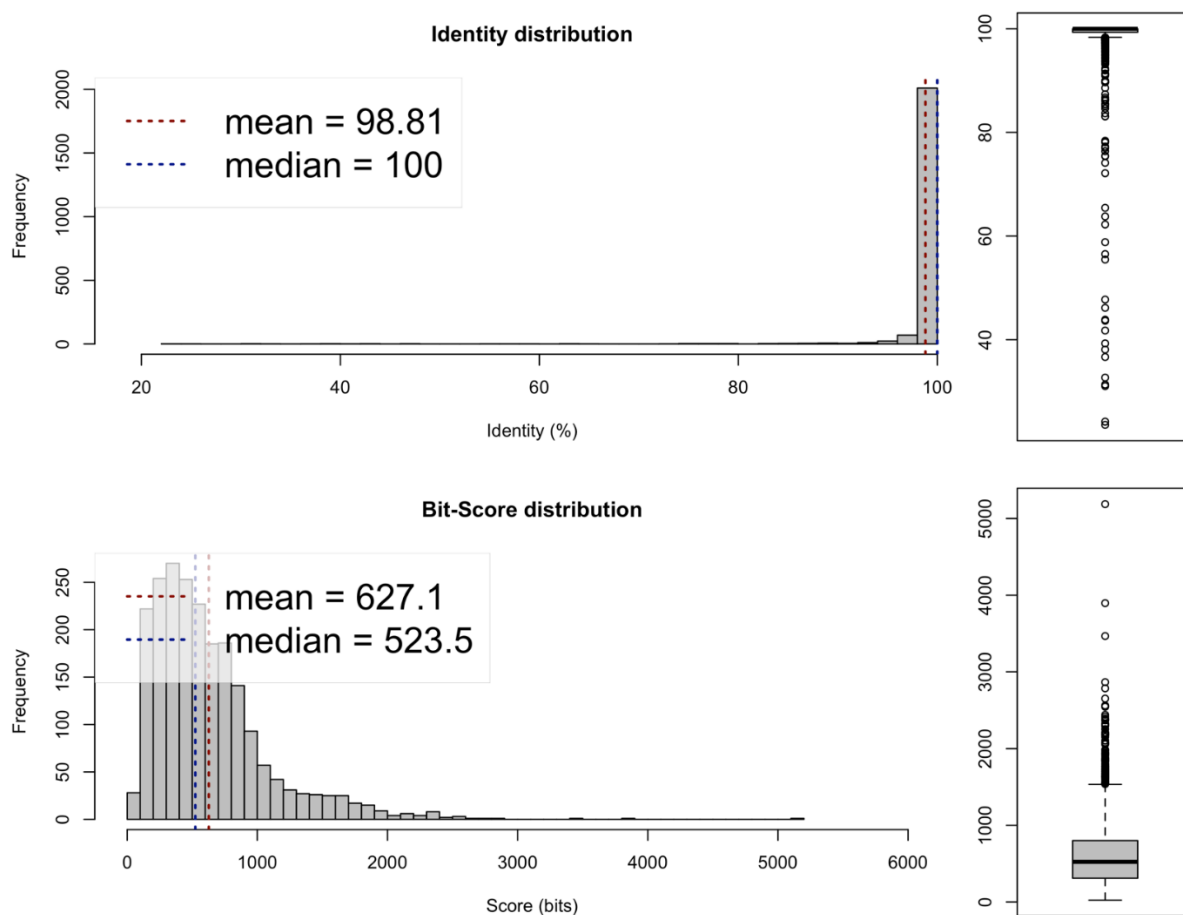


Figure S4-4. Amino acid identity (AAI) and bit score distribution of AOB versus *Nitrosomonas* OLB2

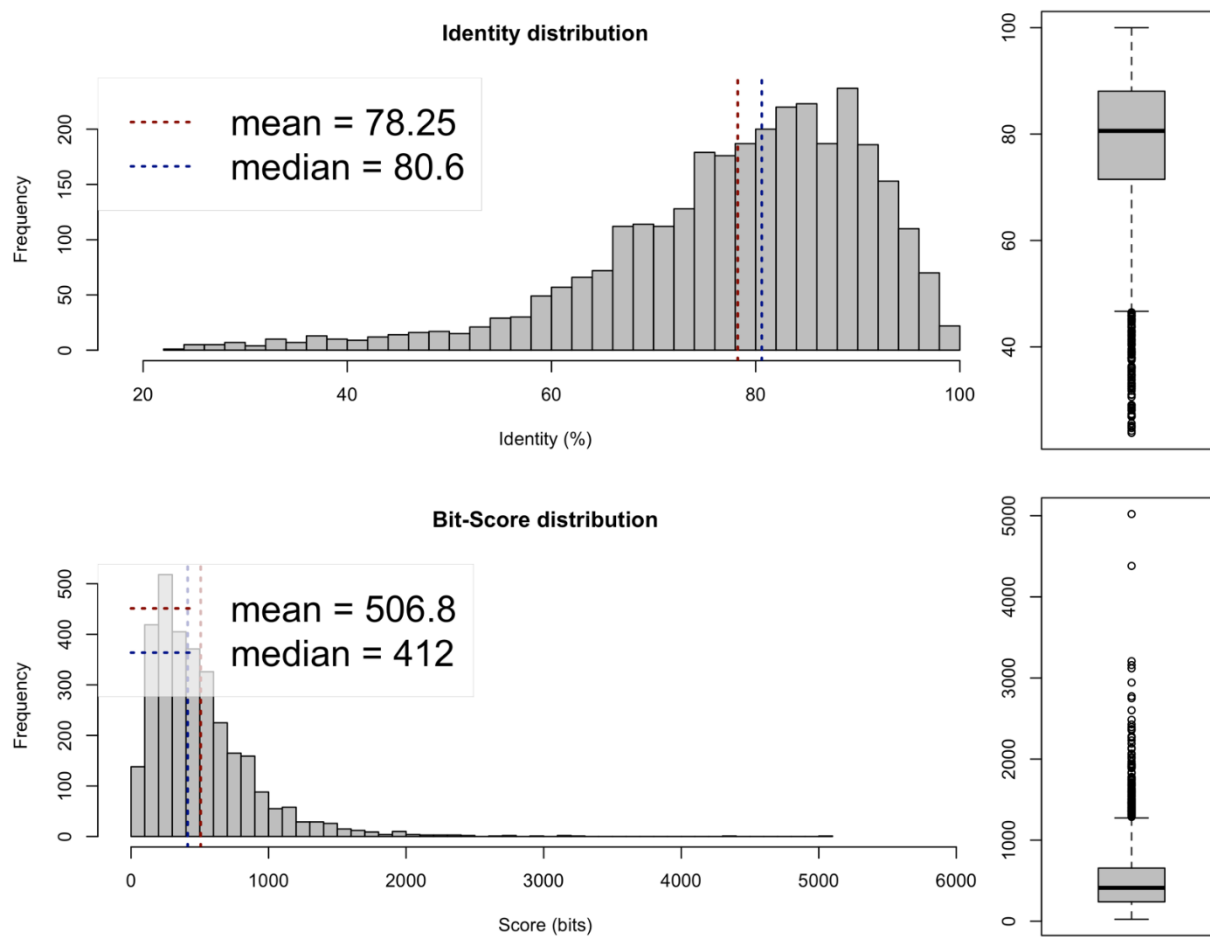


Figure S4-5. Amino acid identity (AAI) and bit score distribution of NOB versus *Nitrospira defluvii*

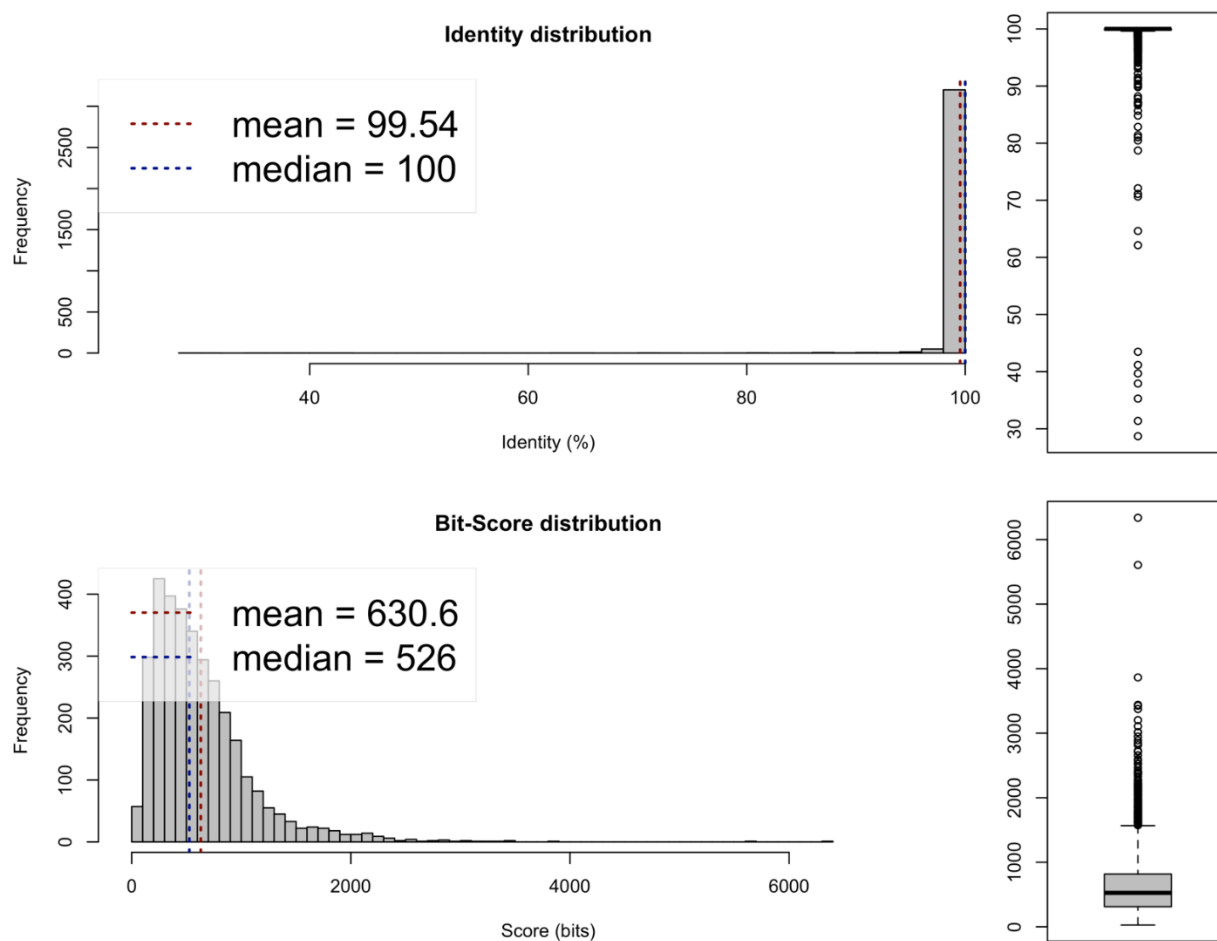


Figure S4-6. Amino acid identity (AAI) and bit score distribution of NOB versus *Nitrospira* OLB3

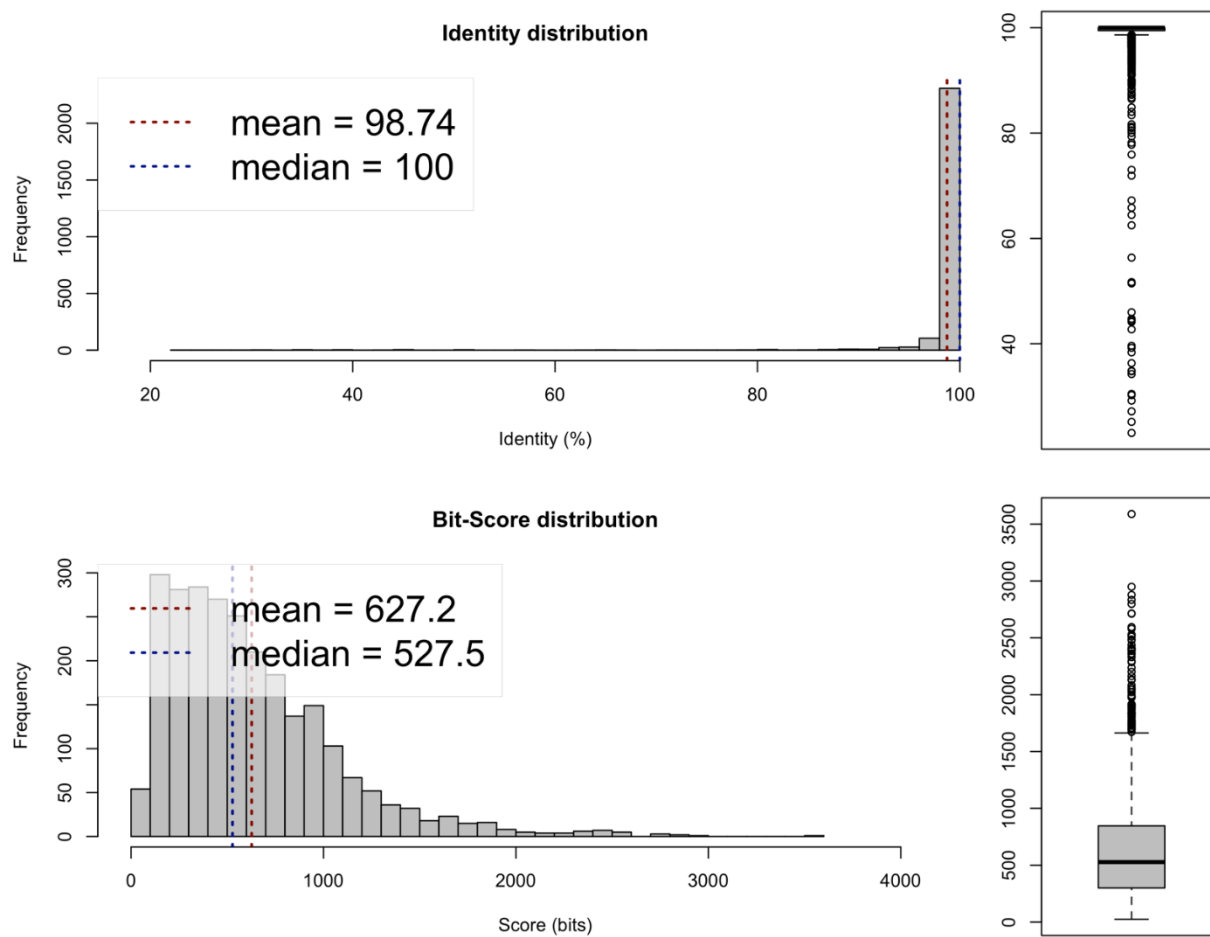


Figure S4-7. Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia fulgida*

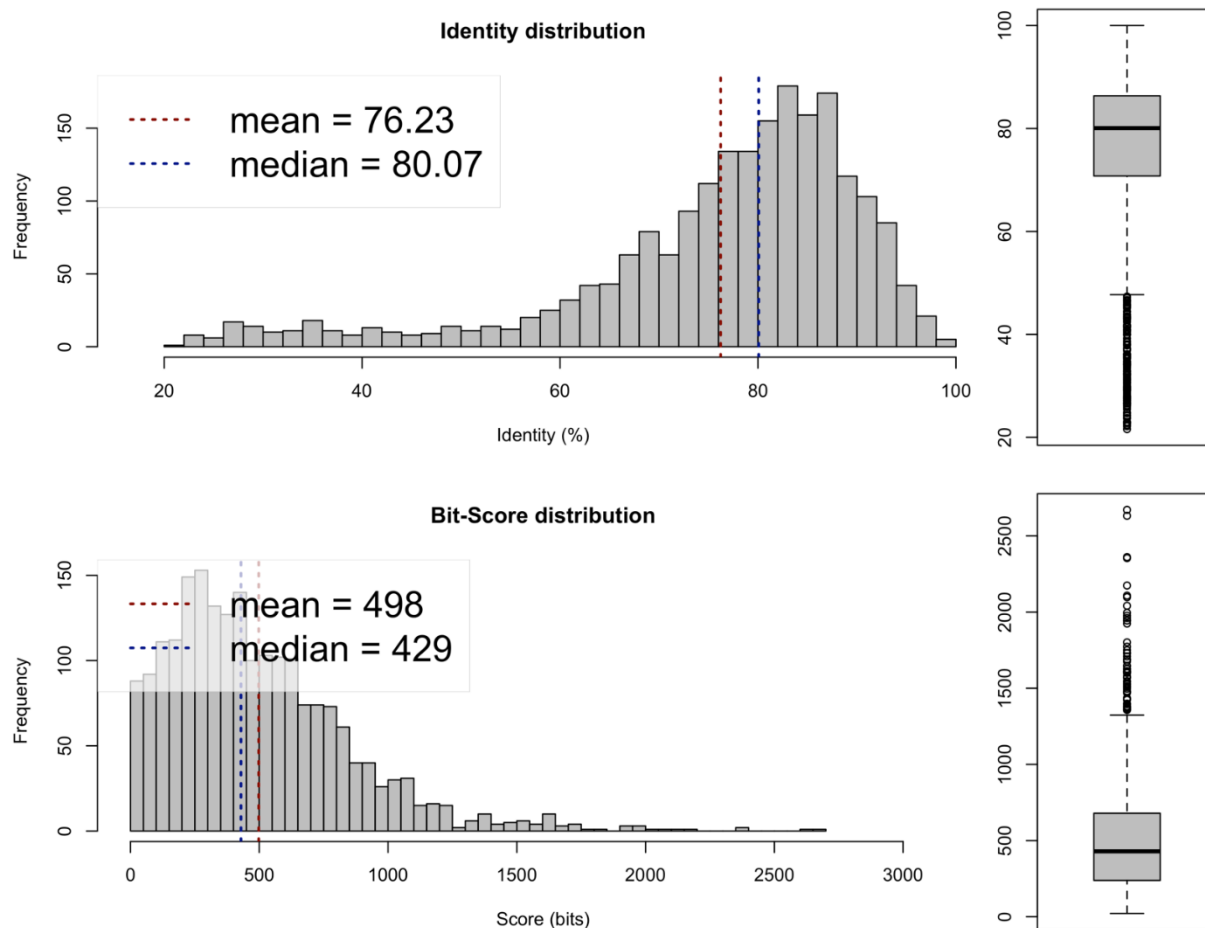


Figure S4-8. Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia sinica* OLB1

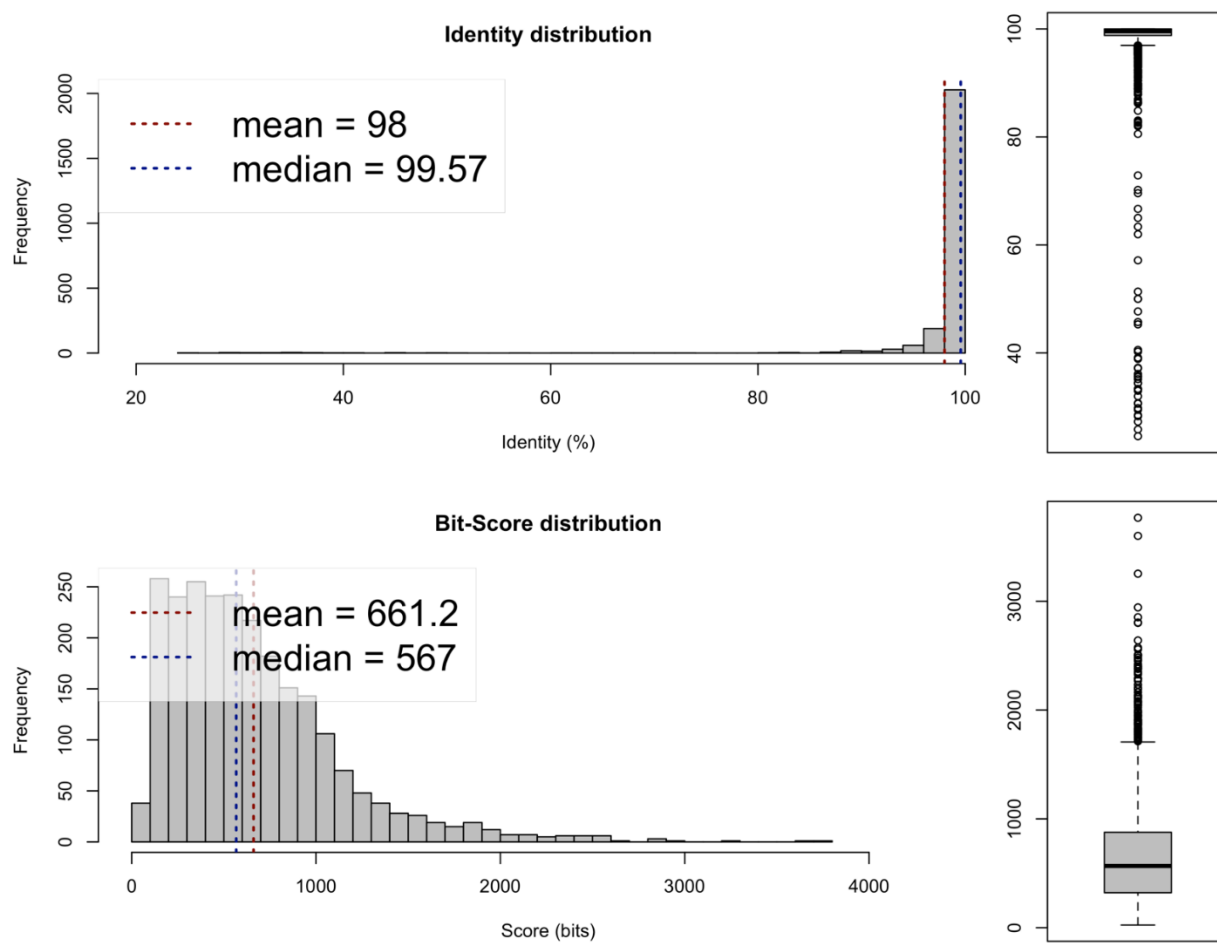


Figure S4-9. Amino acid identity (AAI) and bit score distribution of AMX versus *Ca. Brocadia* UTAMX2

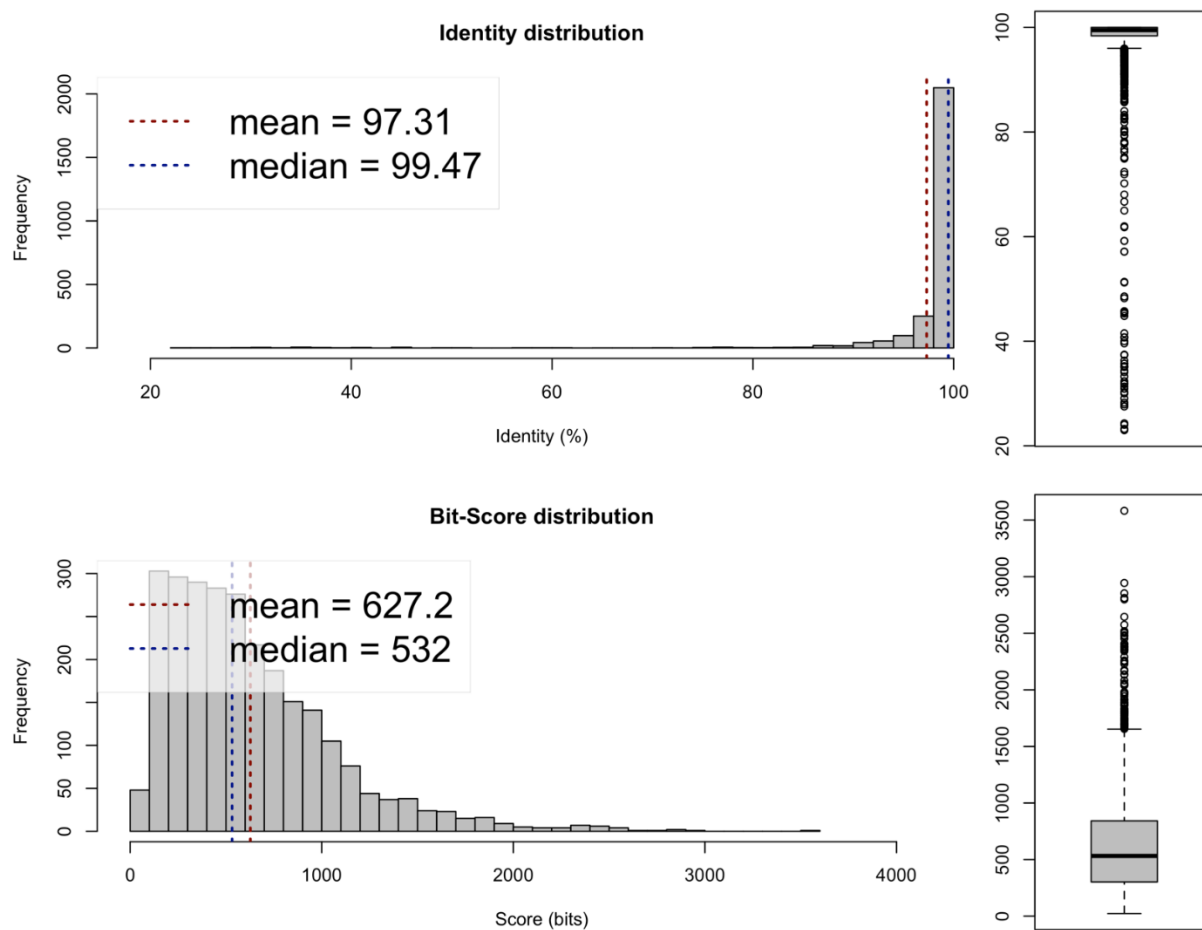


Figure S4-10. Amino acid identity (AAI) and bit score distribution of *Ca. Brocadia* UTAMX2 versus *Ca. Brocadia fulgida*

Chapter 5

Conclusions and Future Directions

Conclusions and Future Directions

In order to work toward the goal of true energy independence at the WRRF facility, it is important to take a wide-angle view and consider the full treatment process (Figure 5-1). This perspective reveals an example of the delicate balance that must be achieved between energy generation, resource recovery, and energy reduction strategies.

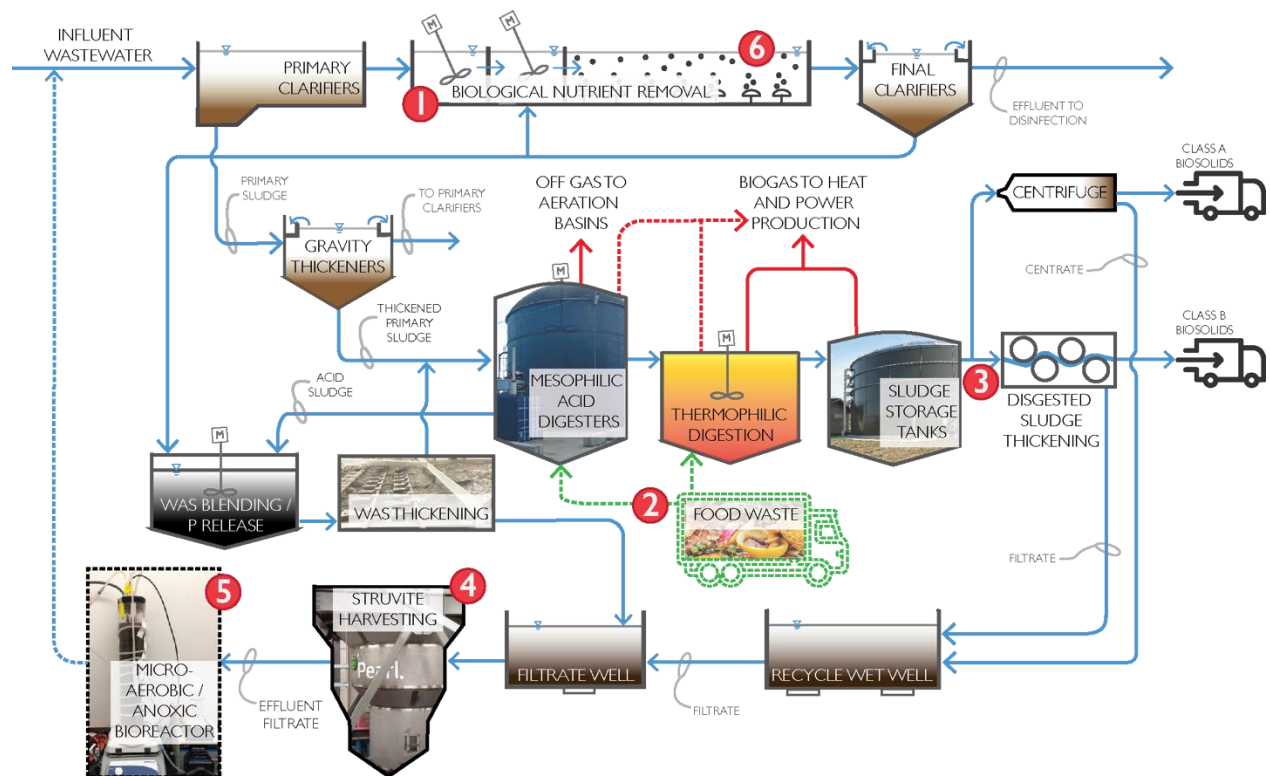


Figure 5-1. WRRF overview with proposed modifications represented as dotted lines. Food waste codigestion is used to boost methane production in the anaerobic digesters (2), while increasing other nutrients in anaerobic digester effluent (3). Phosphorus (P) is released through WAS blending and mixed with digester filtrate and centrate prior to high-quality fertilizer production from struvite harvesting (4). Struvite harvesters precipitate a majority of P (with magnesium and ammonium); however, a high strength N effluent remains and is recycled back to the head of the plant. Since food waste codigestion is expected to increase N loads even further- we propose the addition of a micro-aerobic/anoxic bioreactor (5) following struvite harvesting to alleviate the N load to the mainstream. The simultaneous use of micro-aerobic mainstream BNR (6) reduces overall energy and carbon demand.

Mainstream aerobic biological treatment (Figure 5-1, #1) is one of the most energy intensive processes at the WRRF with aeration, contributing up to 60% of total electricity [3]. In addition, denitrification consumes the limited organic carbon [4] that could be otherwise used for biogas [5, 6] or revenue from medium chain fatty acids production [5]. Thus, treatment strategies that allow us to detach organic carbon and nitrogenous oxygen -demand in nutrient removal, while maintaining effluent goals, is essential.

Research focused on energy generation at the WRRF has promoted food-waste anaerobic codigestion to maximize biogas production for combined heat and power (Figure 5-1, #2) [6, 7]; however, this modification is not without consequences. At surface level, food-waste codigestion appears to be a promising strategy that can ‘kill two birds with one stone’, eliminating waste from ending up in landfills while also furthering WRRF energy security. However, it leaves behind an organic carbon-poor and ammonium-rich wastewater stream (Figure 5-1, #3) that is once again recirculated to the mainstream process creating additional nitrogenous oxygen energy demand [8]— a ‘catch 22’. Some plants have implemented a struvite harvester (Figure 5-1, #4) to precipitate nutrients from recirculating streams, creating an additional revenue from the high-quality fertilizer [9, 10]. Unfortunately, a majority of the N load remains after struvite harvesting [10]. As increasingly stringent discharge limits are underway [11], energy efficiency goals and effluent quality must proceed hand-in-hand.

As an energy saving approach, we show that it is possible to transform a high DO system to low DO operation while maintaining quality effluent. We achieved this by using steady dissolved oxygen changes over a period of 6 months [12]. We also demonstrated

that the microbial community could efficiently removal carbon, nitrogen, and phosphorus long-term with DO concentrations as low as 0.30 O₂/L at the pilot-scale [12]. At low DO, biokinetic testing revealed that microorganisms contributing to phosphorus removal had a high affinity for oxygen in both the full-scale and pilot-scale plants and PAO adaptation to low oxygen conditions was reflected in a greater specific rate of P uptake. We also show that the PAO population structure was unwavering despite DO reductions. On the other hand, biokinetic testing showed that the population structure of microorganisms contributing to ammonia oxidation had shifted toward one with higher affinity for oxygen [12]. Interestingly, the nitrite oxidizing population also remained stable and highly efficient at utilizing oxygen throughout the study [12]. Overall, the low DO pilot plant efficiently removed 99% of the influent ammonia and 89% of the influent phosphorus. Moreover, if the full-scale Nine Springs WRF (Madison, WI) were to follow the same road to low-DO, they could save up to 25% of aeration energy, or approximately \$300,000 per year.

Processes utilizing anammox bacteria for N removal have become extremely popular toward energy-saving wastewater treatment [13-19]—and for good reason, since anammox do not require oxygen or carbon. Micro-aerobic biological N removal (with and without anammox bacteria) has been successfully reported at lab-scale [20-22], pilot-scale [12, 23], and full-scale [24-26]; however, several reports [27-29] have revealed that these systems are highly susceptible to process instability even after stable operation has been reached. We operated a bench-scale anammox bioreactor long-term to treat sidestream reject water from the Nine Springs WRF (Madison, WI). Here we showed that after an initial period of poor performance, reactor performance

appeared to stabilize to ~90% total nitrogen removal. Over 810 days of operation, total inorganic nitrogen load varied between 255 and 1,000 mg-N/day; sCOD varied anywhere between 172 and 6,500 mg O₂/ day and we occasionally observed loss in performance. Sometimes, this loss of performance is easily explained by mechanical issues, such as problems with oxygen delivery. However, in a majority of cases, factors inhibiting microorganisms from participating in N removal are unclear.

A sidestream micro-aerobic/anoxic anammox system (Figure 5-1, #5) together with mainstream micro-aerobic biological nutrient removal (BNR) (Figure 5-1, #6) are an intriguing set of modifications that couple N removal and energy efficiency at a full-scale WRRF. Future studies focusing on genome enabled research gives us the opportunity to not only explore the key microorganisms in micro-aerobic engineered ecosystems, but also allows us to generate hypotheses toward the mechanisms that provide increased fitness in these environments.

References

1. Lebedeva, E.V., et al., *Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon "Ca. Nitrosotenuis uzonensis" representing a clade globally distributed in thermal habitats*. PLoS One, 2013. **8**(11): p. e80835.
2. Henze, M., *Wastewater treatment : biological and chemical processes*. Environmental engineering. 1995, Berlin ; New York: Springer-Verlag. 383 pages.
3. Hamilton, G., et al., *Driving energy efficiency in the US Water & Waste water industry by focusing on operating and maintenance cost reductions*. ACEEE Summer Study on Energy Efficiency in Industry, Washington, DC, 2009.
4. Lu, H., K. Chandran, and D. Stensel, *Microbial ecology of denitrification in biological wastewater treatment*. Water Res, 2014. **64**: p. 237-54.
5. Scarborough, M.J., *Production Of Beneficial Chemicals From Renewable Feedstocks With Anaerobic Microbiomes*. 2018, University of Wisconsin- Madison.
6. Jenicek, P., et al., *Energy self-sufficient sewage wastewater treatment plants: is optimized anaerobic sludge digestion the key?* Water Sci Technol, 2013. **68**(8): p. 1739-44.
7. Zhang, C., et al., *Reviewing the anaerobic digestion of food waste for biogas production*. Renewable and Sustainable Energy Reviews, 2014. **38**: p. 383-392.
8. Nghiem, L.D., et al., *Full scale co-digestion of wastewater sludge and food waste: Bottlenecks and possibilities*. Renewable and Sustainable Energy Reviews, 2017. **72**: p. 354-362.
9. Kataki, S., et al., *Phosphorus recovery as struvite from farm, municipal and industrial waste: Feedstock suitability, methods and pre-treatments*. Waste Management, 2016. **49**: p. 437-454.

10. Mehta, C.M., et al., *Technologies to Recover Nutrients from Waste Streams: A Critical Review*. Critical Reviews in Environmental Science and Technology, 2015. **45**(4): p. 385-427.
11. Hendriks, A. and J.G. Langeveld, *Rethinking Wastewater Treatment Plant Effluent Standards: Nutrient Reduction or Nutrient Control?* Environ Sci Technol, 2017. **51**(9): p. 4735-4737.
12. Keene, N.A., et al., *Pilot plant demonstration of stable and efficient high rate biological nutrient removal with low dissolved oxygen conditions*. Water Res, 2017. **121**: p. 72-85.
13. van de Graaf, A.A., et al., *Anaerobic oxidation of ammonium is a biologically mediated process*. Applied and Environmental Microbiology, 1995. **61**(4): p. 1246-51.
14. Kuenen, J.G., *Anammox bacteria: from discovery to application*. Nat Rev Microbiol, 2008. **6**(4): p. 320-6.
15. Morales, N., et al., *Integration of the Anammox process to the rejection water and main stream lines of WWTPs*. Chemosphere, 2015. **140**: p. 99-105.
16. Wang, S., et al., *Anaerobic ammonium oxidation in traditional municipal wastewater treatment plants with low-strength ammonium loading: Widespread but overlooked*. Water Res, 2015. **84**: p. 66-75.
17. Ma, B., et al., *Biological nitrogen removal from sewage via anammox: Recent advances*. Bioresour Technol, 2016. **200**: p. 981-90.
18. McCarty, P.L., *What is the Best Biological Process for Nitrogen Removal: When and Why?* Environ Sci Technol, 2018. **52**(7): p. 3835-3841.
19. Gao, H., Y.D. Scherson, and G.F. Wells, *Towards energy neutral wastewater treatment: methodology and state of the art*. Environ Sci Process Impacts, 2014. **16**(6): p. 1223-46.

20. Fitzgerald, C.M., et al., *Ammonia-oxidizing microbial communities in reactors with efficient nitrification at low-dissolved oxygen*. Water Res, 2015. **70**: p. 38-51.
21. Park, H.D. and D.R. Noguera, *Evaluating the effect of dissolved oxygen on ammonia-oxidizing bacterial communities in activated sludge*. Water Res, 2004. **38**(14-15): p. 3275-86.
22. Hu, Z., et al., *Nitrogen removal with the anaerobic ammonium oxidation process*. Biotechnol Lett, 2013. **35**(8): p. 1145-54.
23. Camejo, P.Y., et al., *Candidatus Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors*. Water Res, 2016. **102**: p. 125-137.
24. Nifong, A., et al. *Performance of a full-scale sidestream demon installation*. in *Proceedings of the WEF Nutrient Removal and Recovery Conference*. 2013. Vancouver, British Columbia, Canada.
25. Lackner, S., et al., *Full-scale partial nitritation/anammox experiences--an application survey*. Water Res, 2014. **55**: p. 292-303.
26. Park, H.D., et al., *Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors*. Applied and Environmental Microbiology, 2006. **72**(8): p. 5643-5647.
27. Joss, A., et al., *Combined nitritation-anammox: advances in understanding process stability*. Environmental science & technology, 2011. **45**(22): p. 9735-9742.
28. Jin, R.-C., et al., *The inhibition of the Anammox process: A review*. Chemical Engineering Journal, 2012. **197**: p. 67-79.

29. Tang, C.-J., et al., *Thermodynamic and kinetic investigation of anaerobic bioprocesses on ANAMMOX under high organic conditions*. Chemical engineering journal, 2013. **230**: p. 149-157.