bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

High functional diversity among *Nitrospira* populations that dominate rotating biological contactor biofilm communities in a municipal wastewater treatment plant

Emilie Spasov[1], Jackson M. Tsuji[1], Laura A. Hug[1], Andrew C. Doxey[1], Laura A. Sauder[1], Wayne J. Parker[2], Josh D. Neufeld[1#]

[1]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada

[2]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario, Canada

#Corresponding author: Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada. Tel. +1 519-888-4567; Fax +1 519-746-0614. E-mail: jneufeld@uwaterloo.ca

Running title: Comammox *Nitrospira* dominate wastewater biofilm microbial communities

Keywords: complete ammonia oxidation, comammox, ammonia-oxidizing archaea, wastewater, nitrification, biofilm, metagenomics

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Nitrification, the oxidation of ammonia to nitrate via nitrite, is an important process in municipal wastewater treatment plants (WWTPs). Members of the *Nitrospira* genus that contribute to complete ammonia oxidation (comammox) have only recently been discovered and their relevance to engineered water treatment systems is poorly understood. This study investigated distributions of *Nitrospira*, ammonia-oxidizing archaea (AOA), and ammonia-oxidizing bacteria (AOB) in biofilm samples collected from rotating biological contactors (RBCs) of a municipal WWTP in Guelph, Ontario, Canada. Using quantitative PCR (qPCR), 16S rRNA gene sequencing, and metagenomics, our results demonstrate that *Nitrospira* species strongly dominated RBC biofilm samples, and comammox *Nitrospira* outnumber all other nitrifiers. Genome bins recovered from metagenome assemblies reveal multiple populations of comammox *Nitrospira* with distinct spatial and temporal distributions, including several taxa that are distinct from previously characterized *Nitrospira* members. Distinct and diverse functional profiles imply a high level of niche heterogeneity among comammox *Nitrospira*, in contrast to the sole detected AOA representative that was previously cultivated and characterized from the same RBC biofilm. Unexpectedly, our metagenome bins revealed two cyanase-encoding populations of comammox *Nitrospira*, suggesting an ability to degrade cyanate, which has not been shown previously for *Nitrospira* that are not strict nitrite oxidizers. This study demonstrates the importance of RBCs to serve as model systems for continued investigation of environmental factors that control the distributions and activities of AOB, AOA, comammox *Nitrospira*, and other nitrite oxidizers.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Municipal wastewater contains ammonium that is removed by wastewater treatment plants (WWTPs) to prevent eutrophication, oxygen depletion, and toxicity to aquatic animals in receiving waters. Nitrification involves the oxidation of ammonia to nitrate, via nitrite, in two enzymatic steps. These two steps were thought to be mediated by distinct groups of microorganisms, with aerobic ammonia oxidation conducted by ammonia-oxidizing bacteria (AOB) or ammonia-oxidizing archaea (AOA), and nitrite oxidation catalyzed by nitrite-oxidizing bacteria (NOB) [1–3]. The theoretical existence of microorganisms capable of oxidizing both ammonia and nitrite via complete ammonia oxidation (comammox) was predicted over a decade ago, with the suggestion that comammox bacteria would be slow growing and inhabit biofilms that were exposed to relatively low ammonium concentrations [4]. Recently, these predictions were confirmed by the discovery *Nitrospira* members capable of catalyzing comammox [5, 6]. All comammox *Nitrospira* reported to date belong to lineage II of the genus *Nitrospira*. Although two major clades of comammox *Nitrospira* have been described (i.e., clades A and B) based on ammonia monooxygenase (*amoA*) gene phylogeny, all enriched and cultivated species of comammox *Nitrospira* belong to clade A [5, 6].

Compared to AOB and AOA, very little is known about the abundance and diversity of comammox *Nitrospira* in engineered aquatic environments. Consistent with a low ammonium niche, comammox *Nitrospira* have been detected in drinking water systems [7–14]. First identified from water treatment system metagenome sequences [5, 6], most wastewater-associated comammox *Nitrospira* belong to clade A [5, 7, 12, 14–18], albeit with abundances generally lower than those reported for AOA and AOB [7, 18–21]. Nonetheless, *amoA* gene abundances of comammox *Nitrospira* may outnumber those of AOB in activated sludge samples

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in some WWTPs [12, 18, 22], and comammox *Nitrospira* have been enriched in wastewater treatment reactors with low dissolved oxygen conditions, suggesting that comammox bacteria have an important role in ammonia oxidation in these reactors [15, 17]. Additionally, the high abundance of *amoA* transcripts from comammox *Nitrospira* detected in activated sludge suggests that these bacteria actively contribute to nitrification [23]. Previous studies examining comammox *Nitrospira* in WWTPs focused primarily on activated sludge secondary treatment systems and sequencing batch reactors [12, 15–27]. To date, no studies have investigated comammox *Nitrospira* in biofilm-based tertiary wastewater treatment systems.

Rotating biological contactors (RBCs; Figure 1A) represent a full-scale tertiary system treating municipal waste from ~132,000 residents of Guelph, Ontario, Canada. Comprising a total biofilm surface area of 440,000 m^2 and processing ~53,000 m^3 of wastewater daily (Guelph Wastewater Treatment Master Plan, 2009), the RBCs are organized into four "trains" that are each composed of eight individual RBC "stages" (Figure 1B). Continual rotation in and out of post-aeration-basin wastewater allows for alternating exposure of the RBC biofilm to wastewater and air, resulting in decreasing ammonium concentrations along the RBC trains due to aerobic nitrification, as described previously [28]. Because of the relatively low ammonium concentrations entering the RBCs, compared to aeration basins, and their fixed-film design, these systems represent a valuable resource for studying the distribution and diversity of nitrifiers in relation to ammonium concentration gradients. The RBC-associated AOA and AOB communities were characterized previously, demonstrating that the sole detectable AOA population increased in abundance along the RBC flowpath [28]. This AOA species was enriched, characterized, and named *Candidatus* Nitrosocosmicus exaquare [29]. However, inconsistent *in situ* activity data obtained using differential inhibitors for AOB and an abundance

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of *Nitrospira* cells, as viewed by fluorescence *in situ* hybridization (FISH), raised the possibility that some detected *Nitrospira* may contribute to comammox [29], necessitating a re-evaluation of nitrifier communities within the Guelph RBC system biofilms.

Due to the predicted low ammonium niche and biofilm growth of comammox bacteria [4, 5], we hypothesized that comammox *Nitrospira* would dominate RBC biofilm, adding additional complexity to the classical interpretation of RBC microbiology with respect to nitrification. We also predicted that the relative abundance of comammox *Nitrospira* would increase as ammonium concentrations decreased along the RBC flowpath, as demonstrated previously for *Ca.* N. exaquare [28]. To test these hypotheses, we assessed the relative abundance, diversity, and temporal stability of comammox *Nitrospira*, in relation to AOA and other AOB, throughout the tertiary treatment system, using a combination of quantitative PCR (qPCR), 16S rRNA gene sequencing, and metagenomics. Our results provide strong evidence for high functional diversity among populations of Nitrospira within the RBCs, implying high niche heterogeneity and the unexpected potential for use of alternate nitrogen and energy sources by comammox *Nitrospira* in the RBCs, such as cyanate and hydrogen.

Materials and Methods

Sampling

Using ethanol-cleaned spatulas, biofilm was sampled from the RBC trains in Guelph, Ontario, Canada (Figure 1A, Figure 1B). Samples were stored on dry ice until delivered to the lab and were kept at -70°C until analyzed. All RBCs from all four trains were sampled in October 2016, except for RBC 1 of the southeast (SE) train and RBC 2 of the southwest (SW) train, which were not operational at the time of sampling (Table S1). We included original

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

biofilm samples that were collected for a previous study [28] as a temporal comparison. These reference samples were collected in February, June, and September 2010 from RBCs 1 and 8 of the northeast (NE) train (Table S1). Water samples were collected in 2016 from RBC 1 and RBC 8 and kept on ice and then frozen at -20°C until chemical analysis.

Water chemistry

Ammonium was measured fluorometrically using orthophthaldialdehyde (OPA) reagent [30] according to the method by Poulin and Pelletier [31], with minor modifications. Volumes of 100 µL of sample and 200 µL of OPA working reagent were added to a 96-well opaque flat-bottomed plate. Plates were incubated for four hours in the dark before being measured. Nitrite and nitrate concentrations were assessed colorimetrically using the Greiss reagent as described elsewhere [32]. All samples were measured as technical duplicates at 360 nm excitation, 465 nm emission (ammonium), and 550 nm (nitrite/nitrate) using a FilterMax F5 Multi-Mode Microplate reader (Molecular Devices, San Jose, CA, USA). Measurements represent total ammonium as nitrogen, and nitrite and nitrate as nitrogen. Based on data obtained from WWTP operators, monthly and yearly ammonium concentrations (measured as total ammonia as nitrogen) of secondary effluent (i.e., RBC influent) and water from the southeast RBC train were compiled for the years 2010 to 2017. Plant influent and effluent data were retrieved from the Guelph WWTP annual reports for 2010 to 2017 (https://guelph.ca/living/environment/water/wastewater/). The pH of water samples was measured with a DELTA 320 pH meter (Mettler Toledo, Mississauga, ON, Canada).

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DNA extractions

All DNA extractions were performed from 0.25 g (wet weight) of biofilm with the PowerSoil DNA Isolation Kit (Mo Bio, Carlsbad, CA, USA), as described previously [29]. Total isolated DNA was visualized on a 1% agarose gel and quantified using a NanoDrop 2000 (Thermo Scientific, Waltham, MA, USA) and Qubit dsDNA high sensitivity assay kit (Thermo Scientific).

Extracted DNA from RBC 1 and 8 biofilm samples of the NE train from 2010 was used for quantitative PCR (qPCR) and metagenome sequencing (Table S1). The 2010 biofilm samples were those collected and analyzed by qPCR previously [28], but DNA extractions were repeated while processing the 2016 samples. The DNA extracts from all RBCs within the four trains sampled in 2016 were analyzed with qPCR, whereas DNA from RBCs 1 and 8 alone were used for 16S rRNA gene and metagenome sequencing (Table S1).

Quantitative PCR

Thaumarcheotal and bacterial 16S rRNA genes were quantified using the primers 771F/957R [33] and 341F/518R [34], respectively (Table S2). Quantification of AOB *amoA* genes was carried out using the primers amoA1F/amoA2R [35] (Table S2). Sequences from comammox *Nitrospira* clade A and clade B *amoA* genes were amplified using equimolar primer mixes of comaA-244f (a-f) and comaA-659r (a-f), and comaB-244f (a-f) and comaB-659r (a-f), respectively [7] (Table S2). Primers for each of comammox *Nitrospira* clades A and B *amoA* genes were initially tested in end-point PCR to check for a single dominant band in all amplifications, but subsequent qPCR was performed with clade A primers only because clade B primers produced no amplicons (data not shown). All qPCR amplifications were carried out as

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

technical duplicates on a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). Additional details on thermal cycling conditions and standards are included in the supplemental methods.

16S rRNA amplicon sequencing

The V4-V5 region of the 16S rRNA gene was amplified by PCR using the primers 515F [36] and 926R [37] with Illumina adapters. The sequence data were produced by the US Department of Energy Joint Genome Institute using standard operating procedures. Triplicates were combined, quantified, and sample amplicons were then pooled by equal amounts. The pooled library was sequenced on a MiSeq (Illumina) with a 2 x 300 bp run. Sequence analysis was performed with QIIME2 version 2019.1.0 [38]. Sequences were trimmed to remove primer and adapter sequences with cutadapt [39] within QIIME2. Quality trimming, denoising, error-correction, paired-end read merging, chimera removal, and dereplication was performed with DADA2 [40], within QIIME2. This produced an amplicon sequence variant (ASV) table for analysis and 80,180-125,219 assembled sequences per sample (Supplemental file 1). The ASVs were taxonomically classified according to the SILVA database release 132 [41] using the scikit-learn classifier [42] within QIIME2. The unrarefied abundances of the ASVs were plotted as proportional abundances using a custom R script.

Metagenome sequencing and analysis

Shearing of DNA, library preparation, and sequencing were performed at The Centre for Applied Genomics (TCAG) in Toronto, Ontario. Extracted DNA was quantified by Qubit dsDNA High Sensitivity assay (Thermo Scientific, Waltham, MA, USA) and 500 ng of input DNA was sheared into ~550 bp fragments using a Covaris LE220 Focused-ultrasonicator

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Covaris, Woburn, MA, USA). Library preparation was done with the TruSeq PCR-free Library Prep Kit (Illumina, San Diego, CA, USA). Paired-end sequencing (2x250 bases) was performed on a HiSeq 2500 (Illumina) using the HiSeq Rapid SBS Kit v2 (500 cycle) (Illumina), resulting in a total of ~250 million paired-end reads with an average of 18.1 million paired-end reads per sample (Table S3).

Quality trimming and removal of adaptor sequences was performed using AdapterRemoval version 2.2.2 [43], and the quality of the reads was checked with FastQC version 0.11.5 [44]. Reports were combined using MultiQC version 1.0 [45]. Open reading frames (ORFs) were predicted on the unmerged and unassembled forward trimmed reads using FragGeneScan-Plus [46]. Profile hidden Markov models (HMMs) for taxonomic marker (*rpoB*) and functional genes (*amoA*_AOA, *amoA*_AOB, and *nxrB*), downloaded from FunGene [47], were used to quantify the relative abundances and taxonomic affiliations of nitrifiers from the unassembled reads using MetAnnotate [48]. The HMM e-value threshold used by MetAnnotate for gene detection was 10^{-3}, and the e-value threshold used for the USEARCH-based taxonomic classification step was 10^{-6}. Separate *amoA* HMMs for AOA and AOB were used for improved gene recovery. Within MetAnnotate, the database used for taxonomic classification was RefSeq release 80 (March 2017). MetAnnotate results were analyzed using the custom R script *metannotate_barplots.R* version 0.9 (available at https://github.com/jmtsuji/metannotate-analysis). To allow for approximate between-sample and between-HMM comparisons, hit counts for each HMM were normalized both to the HMM length and to the total number of length-normalized HMM hits for *rpoB* (for more details, see the Github README).

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Assembly and binning of metagenome sequence reads

Raw metagenome sequence reads were additionally processed through the ATLAS pipeline (version 2.0.6), which includes quality control, assembly, annotation, binning, and read mapping of metagenome sequence data [49, 50]. For further details, see supplemental methods. Taxonomy was assigned to bins according to the Genome Taxonomy database [51] release 86, version 3, using the Genome Tree Database Toolkit version 0.2.2 (GTDB-Tk https://github.com/Ecogenomics/GtdbTk). Relative abundances of genome bins in metagenomic data were approximated by calculating the number of mapped reads to genome bins normalized to the total number of assembled reads per metagenome. Beta diversity of the samples was determined based on these relative abundance data. Bray-Curtis dissimilarities and the principal coordinate analysis (PCoA) ordination were calculated using a custom R script.

Analysis of genome bins

Average nucleotide identity (ANI) was calculated for the dereplicated AOA bin and *Nitrospira* bins compared to reference genomes using FastANI version 1.1 [52]. The *Nitrospira* bins were further searched for genes involved in nitrogen cycling using reciprocal BLASTP [53] of known genes in three reference *Nitrospira* genomes (*Ca.* N. inopinata, *Nitrospira moscoviensis*, and *Nitrospira lenta*). The searches were performed using the BackBLAST pipeline [54], development version 2.0.0-alpha (https://github.com/LeeBergstrand/BackBLAST_Reciprocal_BLAST), and the e-value threshold and identity cutoff were manually optimized to 1e-30 and 40%, respectively, based on expected hits to reference genomes. To prevent the incorrect exclusion of true orthologs among AmoC proteins, the AmoC3 protein sequence (WP_062483313.1) was removed from the reference

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proteome of *Ca.* N. inopinata during the BLASTP search due to its close sequence identity to the query AmoC1 protein (WP_062484140.1), which interfered with reciprocal BLASTP.

A concatenated core protein phylogeny was created to assess the phylogenetic placement of the *Nitrospira* genome bins. Reference genomes were downloaded from the NCBI database and a core set of 74 core bacterial proteins ('Bacteria.hmm' in GToTree) were identified, aligned, concatenated, and used for phylogenetic tree construction from these and the *Nitrospira* genome bins via the GToTree pipeline, version 1.1.10 [55], with default settings. The maximum likelihood phylogeny was computed using IQ-TREE version 1.6.9 [56] with 1000 bootstrap replicates and the LG+F+R10 evolutionary model as selected by IQ-TREE's ModelFinder module [57] .

Functional gene phylogenies

All predicted protein sequences from assembled contigs were compiled and clustered into 99% identity threshold groups within ATLAS. This clustered sequence set was searched with the "amoA_AOB" profile hidden Markov model (HMM) from the FunGene database [47] using hmmsearch version 3.1b2 [58] with an e-value cutoff of 1e-10 to identify putative AmoA sequences. Identified hits were queried against the NCBI RefSeq database via BLASTP [53] and those with closest database matches to the *Nitrospirota* phylum were used for downstream analysis. Using MUSCLE [59] within MEGA7 [60], sequences of identified proteins were aligned, along with other AmoA sequences from enrichment cultures and environmental surveys retrieved from the NCBI database. Phylogenetic analysis was inferred using maximum likelihood

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis with 500 bootstrap replicates, based on the Le Gascuel evolutionary model [61], using MEGA7.

Cyanase protein sequences (CynS) identified via BackBLAST were further analyzed phylogenetically. Reference annotated CynS sequences were identified from the SwissProt database and downloaded from RefSeq for consistency with other analytical steps. These references were aligned to the CynS sequences identified in *Nitrospira* genomes using Clustal Omega version 1.2.3 [62, 63], and the alignment was used to build a maximum likelihood phylogeny using IQ-TREE version 1.6.11 with 1000 rapid bootstraps [56]. The LG+I+G4 evolutionary model was selected based on the ModelFinder module of IQ-TREE [57]. The two annotated archaeal reference sequences of CynS were chosen as the outgroup for the phylogeny as shown previously [64]. Subsequently, all genome bins in the metagenomic data were searched for the presence of CynS using hmmsearch version 3.1b2 [58]. The Pfam profile HMM PF02560, which covers the unique C-terminal domain of CynS, was used to perform the search with an e-value cutoff of 1e-20. Read mapping statistics for genome bins containing CynS, along with the CynS multiple sequence alignment and phylogeny, were plotted using a custom R script.

Data availability

The amplicon sequence data are available on the JGI genome portal under sequencing project ID 1137811. Metagenome sequencing data were deposited in the European Nucleotide Archive under study accession number PRJEB30654.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Wastewater nitrogen concentrations

Ammonium concentrations were higher in RBC 1 (>9 μM) than RBC 8 (<4 μM) for all treatment trains (Figure 1C, Table S4) sampled in 2016 and, likewise, nitrite concentrations were greater in RBC 1 (>18 μM) than RBC 8 (<10 μM) for most sampled RBCs. In contrast, nitrate concentrations (>350 μM in all RBCs) greatly exceeded ammonium and nitrite (Figure 1C, Table S4). Although average plant influent and effluent ammonium concentrations were relatively stable from 2010 to 2016 (Figure S1A), data obtained from plant operators demonstrate that average ammonium concentrations of the RBC influent were lowest in 2012 (16.4 μM) and 2016 (22.1 μM) (Figure S1B). Average ammonium RBC influent concentration in 2016 (22.1 μM) was lower than in 2011 (81.4 μM), the difference was not significant (independent samples t-test, $p = 0.07$).

Gene abundances in biofilm samples

Thaumarchaeotal 16S rRNA genes, *amoA* genes from AOB, and comammox *Nitrospira* clade A *amoA* genes were detected in all samples by qPCR (Figure 2, Figure S2); no clade B comammox *Nitrospira amoA* genes were detected in any samples (data not shown). The qPCR data imply distinct nitrifying community compositions, both along the treatment train and between sampling time points. Within each RBC train, the number of comammox *Nitrospira amoA* genes, as well as the proportion of comammox *Nitrospira* to the total community, were consistently higher in RBC 1 than RBC 8 (Figure 2, Table S2). For the 2010 samples, there was a higher abundance of thaumarchaeotal 16S rRNA genes in RBC 8 than RBC 1; the proportion of AOA to the total community was highest in RBC 8 samples. In contrast, thaumarchaeotal 16S

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rRNA gene abundances were higher in RBC 1 than RBC 8 in all four trains sampled in 2016 and were roughly equivalent in abundance to comammox *Nitrospira* (Figure 2). Although the abundances of AOB *amoA* genes were higher in RBC 1 than RBC 8 in both sampling years, the abundances of these genes were one to two orders of magnitude lower in 2016 than in 2010.

Comammox *Nitrospira amoA* genes were more abundant than those of other ammonia oxidizers for most samples (Figure 2, Figure S2, Table S5). In 2010, comammox *Nitrospira amoA* genes outnumbered AOA 16S rRNA and AOB *amoA* genes for all samples, except for the September RBC 8 sample where AOA genes outnumbered genes of other ammonia oxidizers. For these 2010 samples, AOB *amoA* genes outnumbered AOA 16S rRNA genes in RBC 1 at all timepoints, whereas AOA genes either outnumbered AOB genes in RBC 8, or their abundances were roughly equal. In 2016, comammox *Nitrospira amoA* genes outnumbered the genes of the other ammonia oxidizers, except in NE8 and NW8, where AOA genes outnumbered comammox *Nitrospira* genes. The AOA genes greatly outnumbered AOB genes in RBCs 1 and 8 for all trains sampled in 2016. Higher resolution sampling of all 2016 RBC stages showed the general trend of a decrease in *amoA* gene abundances across all four RBC flowpaths (Figure S2).

Dominant nitrifiers detected by 16S rRNA gene sequencing

For all RBC samples that were subsequently analyzed by metagenomics, we generated 16S rRNA gene profiles (Supplemental file 1) to complement qPCR data. The results demonstrate that eight amplicon sequence variants (ASVs) present at ≥1% relative abundance were classified as *Nitrospira* (Figure 3) and, taken together, these ASVs dominated all other microbial community members detected in the RBC biofilm samples. Although different relative abundance patterns for each *Nitrospira* ASV were observed (Figure 3), the overall pattern of relative abundances for all *Nitrospira* together matched patterns seen with qPCR data (Figure 2),

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with higher abundances of *Nitrospira* detected in RBC 1 compared to RBC 8 for each sample pair. Because the 16S rRNA gene cannot be used to differentiate between comammox *Nitrospira* and strict nitrite oxidizers [7], these abundances would potentially reflect both cohorts. Across all samples, five ASVs present at ≥1% relative abundance were classified into the *Nitrosomonadaceae* (Figure 3), with four corresponding to the genus *Nitrosomonas* and affiliated only with 2010 samples. Although one *Nitrosomonadaceae* ASV affiliated with 2016 samples (ASV ID 4), subsequent metagenome analysis suggested that this was not an ammonia oxidizer due to lack of a detectable *amo* gene system in the corresponding genome bin (data not shown). Thus, qPCR data were consistent with ASV data, with fewer AOB detected in 2016 samples overall. We detected one ASV affiliated with the known AOA population within the RBCs (*Nitrocosmicus* genus) and this ASV was present in all samples (Supplemental file 1) but at highest relative abundance (i.e., >1%) for 2016 samples (Figure 3), as expected based on qPCR data (Figure 2). The overall 16S rRNA gene profiles grouped by year and RBC number (Figure S3A), demonstrating that distinct microbial communities were present in each of the two sampling years and along the RBC flowpaths.

Biofilm community composition from metagenomics

Based on single copy taxonomic marker gene analysis of unassembled metagenome sequence data, *Nitrospira* sequences dominated all samples and comprised between 8 and 32% of the total *rpoB* sequences among RBC samples from the two sampling years (Figure 4). The *Nitrospira*-affiliated *rpoB* gene relative abundances were higher in RBC 1 than RBC 8 for each train sampled. Consistent with qPCR data, *Nitrosomonas*-affiliated *rpoB* gene sequences were detected at ≥1% in the 2010 samples only. The community composition of dominant microorganisms was stable between seasonal timepoints in 2010 and these profiles were distinct

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from those detected in 2016. The main difference was the decrease in *Nitrosomonas rpoB* gene abundances in 2016, making *Nitrospira* the only potential ammonia oxidizer detected at ≥1% for this later sampling time point. The *rpoB* gene of AOA was not detected at ≥1% relative abundance in any of the samples, but ranged from 0.01-0.3% across all samples (data not shown). Due to the decreased specificity of taxonomic assignment at the species level [48], we analyzed taxonomic classification at the genus level and, as such, taxonomic markers such as *rpoB* cannot indicate whether these *Nitrospira* spp. were strict nitrite oxidizers or were *amoA*-containing comammox bacteria.

The HMM search for the *amoA* gene of AOA revealed low relative abundance in the metagenome dataset (<2% relative to *rpoB* for all samples; Figure 5A). For 2010 samples, *amoA* gene sequences affiliated with *Thaumarchaeota* were more abundant in RBC 8 than RBC 1 samples, whereas in 2016 they were more abundant in RBC 1 than RBC 8 samples, except in the NE 2016 sample. The HMM hits for AOA *amoA* genes in the unassembled metagenome sequences were low, with only 64 hits to this HMM for all samples combined. Despite this, the pattern of *amoA* gene sequence abundance (normalized to *rpoB*) was correlated with the corresponding proportion of AOA 16S rRNA genes within the total community determined via qPCR (Spearman's rank correlation, $r_s = 0.67$, $p = 0.009$).

The *amoA* HMM for AOB detected sequences of both comammox *Nitrospira* and AOB. Overall, the relative abundances of AOB *amoA* genes were higher in RBC 1 than RBC 8 of the corresponding sample pairs, in both sampling years (Figure 5B). In 2010, *Nitrosomonas* spp. and *Nitrospira* spp. were at roughly equal relative abundances, whereas *Nitrospira* spp. were the dominant ammonia oxidizers detected in 2016 metagenome sequences. The *amoA* genes of *Nitrosospira* were also detected, but always at <1% relative abundance. Assuming one copy of

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

amoA per genome, comammox *Nitrospira* represented between ~5-30% of the total microbial community for all samples based on normalized *amoA* HMM hits (Figure 5B). In addition, *Nitrospira*-associated *rpoB* and normalized *amoA* gene abundances were approximately equal (Figure 4, Figure 5B), suggesting that most *Nitrospira* present in the RBCs were comammox bacteria. Overall, the relative abundance patterns for AOB and comammox *Nitrospira amoA* genes correlated well with their corresponding relative abundances determined by qPCR for the proportion of their *amoA* genes within the total community (AOB $r_s = 0.97$, $p < 0.001$; comammox *Nitrospira* $r_s = 0.79$, $p < 0.001$).

The HMM search for *nxrB* revealed that sequences affiliated with this model were prevalent in RBC biofilm metagenomes, with combined hits at over 50% abundance normalized to total *rpoB* hits (Figure 5C). Although this indicates a high relative abundance of nitrite oxidizers, taxonomic affiliations of some hits suggest false positives beyond *Nitrospira* spp. affiliations. The high relative abundance of *Nitrospira*-affiliated *nxrB* gene sequences, which were more abundant in RBC 1 than RBC 8 metagenome sequences across all samples (Figure 5C), is further evidence for *Nitrospira* dominance in the RBCs. Normalized relative abundances of *Nitrospira*-affiliated *nxrB* sequences that exceed the corresponding relative abundances of *Nitrospira*-affiliated *rpoB* sequences may be due to multiple copies of *nxrB* common to *Nitrospira* genomes [65], in contrast to single copies common for *rpoB* genes [66].

Functional diversity with genome bins from metagenome sequences

The final dereplicated genome bin set included a total of 101 bins (Supplementary file 2). Of this, 15 were considered *Nitrospira* metagenome assembled genomes (MAGs), with high completeness (≥75%) and low contamination (<10 %) (Table 1). These *Nitrospira* MAGs were

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

370 classified under the Genome Taxonomy Database (GTDB) taxonomy as either *Nitrospira*, *Nitrospira* A, or UBA8639 genera, all under the order *Nitrospirales*. Using a concatenated set of 74 core bacterial proteins, these MAGs were further classified according to traditional *Nitrospira* phylogeny (Figure 6). A majority (10) of the *Nitrospira* MAGs clustered within clade A comammox *Nitrospira*, four clustered within sublineage I (strict nitrite-oxidizing bacteria) and 375 one clustered within sublineage IV (strict nitrite-oxidizing bacteria). Using average nucleotide identity (ANI) analysis, the 15 *Nitrospira* MAGs were compared to each other, and to *Nitrospira* genomes from enrichment cultures and metagenomic surveys (Figure S4). Only one MAG, RBC035, had ≥95% ANI to previously recovered comammox genomes (*Nitrospira* bin UW_LDO_01 [26] and *Candidatus* Nitrospira nitrosa [6]); all other comammox *Nitrospira* 380 MAGs had <95% ANI to other reference genomes. This suggests that these MAGs represent distinct comammox *Nitrospira* species from most other previously described comammox species.

Using a reciprocal BLAST analysis, the *Nitrospira* MAGs were further evaluated for the presence of *amo*, *hao*, and *nxr* genes to explore their potential to perform comammox [5, 6]. 385 *Nitrospira* bin RBC047 contained all the genes necessary to perform complete ammonia oxidation, though it was missing an ammonium transporter gene (Figure 6). *Nitrospira* bins RBC069, RBC001, and RBC044 were missing *nxrB* genes, but contained the other genes for ammonia and nitrite oxidation and had a Rh50 type ammonium transporter. In total, there were five clade A MAGs that contained the *amoA* gene, which is used as a marker gene for 390 comammox *Nitrospira* (see below). The other five MAGs that clustered phylogenetically within clade A comammox *Nitrospira* contained other genes for ammonia oxidation, supporting their phylogenetic placement on the tree. The other *Nitrospira* MAGs contained genes only expected

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in NOB, except bin RBC026, which had genes for hydroxylamine oxidation. The high (93%) ANI between this bin and comammox bin RBC083 suggests some level of contamination with comammox sequences is present in bin RBC026. Overall, one *Nitrospira* MAG contained the entire gene pathway required for comammox, four bins contained only genes expected for NOB, and nine bin clusters could potentially represent comammox *Nitrospira* but lacked the complete gene pathway.

The comammox *Nitrospira* identified in this study likely have the ability to hydrolyze urea, given a nearly complete *ure* operon on contigs from most clade A comammox *Nitrospira* MAGs (Figure 6). Although these *amo*-containing *Nitrospira* MAGs did not contain genes for formate dehydrogenase (*fdh*), most MAGs were associated with annotations for a group 3 [Ni-Fe] sulfur-reducing hydrogenase (*hyb* and *hyd* genes). Several of these MAGs also contained the genes for a putative type hydrogenase (*hyf*). Multiple clade A comammox MAGs contained genes for accessory proteins for hydrogenases (*hyp*) but none of the *Nitrospira* MAGs contained group 2a [Ni-Fe] hydrogenase genes (*hup*). All of the comammox *Nitrospira* MAGs lacked the gene for assimilatory nitrite reduction (*nirA*).

The sublineage I MAGs contained genes for cyanate hydratase (*cynS*) and most of these also contained genes for formate oxidation (*fdh*). Surprisingly, two clade A comammox Nitrospira bins (RBC069 and RBC093) that contained genes for ammonia oxidation also contained the *cynS* gene. These genes were on long contigs (182 kb and 121 kb, RBC069 and RBC093, respectively), containing other genes that were classified as belonging to comammox *Nitrospira*. In addition, in RBC093, there were two toxin-antitoxin genes in the vicinity of the cyanase gene. One cyanate transporter gene (*cynD*) seemed to have moderate homology to another non-orthologous ABC transport gene, as indicated by the low amino acid identity hits to

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cynD among certain *Nitrospira*. These hits may represent another ABC transport protein, a nitrate/sulfonate/bicarbonate ABC transporter (accession number WP_121988621). The cyanase (*cynS*) genes were also distinct from those of other strict NOB *Nitrospira*, sharing <75% identity (blastn) to cyanase genes of described *Nitrospira* species in the non-redundant nucleotide database. The cyanase genes in the bins of sublineage I *Nitrospira* shared >86% identity with *Nitrospira defluvii* cyanase gene. Further characterization indicated that the cyanases in the comammox *Nitrospira* bins are basal to previously characterized NOB *Nitrospira* (Figure 7A). The protein sequences contain the active site residues (Arg96, Glu99 and Ser122) for cyanase, as proposed previously [67]. Bins containing *cynS* together represented a substantial proportion of the RBC microbial community, with *Nitrospira* bins dominating the collection of cyanase-containing bins (Figure 7B).

The *Nitrospira* MAGs displayed distinct temporal and spatial distributions. Summed together, the *Nitrospira* MAGs had a higher relative abundance (measured as total recruited mapped reads) in RBC 1 than RBC 8 for each sample pair (Figure 8). Taken as a group, this pattern was also the case for the bins classified as clade A comammox *Nitrospira*. These comammox MAGs comprised a majority of the total recruited reads for all *Nitrospira* in most samples. Some of the comammox MAGs had different abundance patterns in 2010 than in 2016; RBC035, RBC100, and RBC069 were present at higher relative abundance in 2010 samples than in 2016 samples, and relative abundances of RBC001 and RBC083 were generally higher in 2016 than 2010 samples (Figure 8, Supplementary file 2). Multiple comammox *Nitrospira* populations were found in the same RBC and this high diversity of comammox *Nitrospira* contrasts with low AOA diversity (i.e., only one *Nitrosocosmicus* species; see below) for all RBCs sampled. Using the all of the dereplicated bins, the overall microbial communities of the

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

samples group by year and RBC number (Figure S3B), in a similar pattern to that observed with 16S rRNA sequence data (Figure S3A).

Genome bins of other ammonia oxidizers were also recovered. Following dereplication, four bins were classified as *Nitrosomonas* (Supplementary file 2). These populations were present at a higher relative abundance in 2010 samples than 2016 samples. A single dereplicated *Nitrosocosmicus* (AOA) bin was detected (RBC071; Supplementary file 2), which indicates the presence of a single *Thaumarchaeota* population and matches previous data from the RBCs [28, 29]. In 2010, this species was present at a higher relative abundance in RBC 8 than RBC 1, with the opposite pattern occurring in 2016 samples. It was classified as *Ca.* N. exaquare under GTDB taxonomy, and had an ANI of 96.1% between it and the *Ca.* N. exaquare genome (GenBank accession number CP017922.1), which indicated that the AOA MAG corresponded to the same archaeon enriched and characterized from a RBC biofilm sample.

Nitrospira *phylogenetic relationships based on* amoA

From all assembled contigs, there was a total of 20 unique *amoA* gene sequences, based on 99% similarity clustering of sequences of amino acids (data not shown). Of these, ten were classified as *Nitrospira* and three of these unique sequences ended up in the final bin set (Figure S5). Six *amoA* sequences were classified as AOB (*Nitrosomonas* or *Nitrosospira*), and one was classified as an AOA (*Nitrosocosmicus*; data not shown). The other *amoA* sequences were classified as *Bradyrhizobium* and *Rhodoferax*, but these sequences did not end up in bins for the final dereplicated bin set (data not shown). Aside from RBC group F (which was a truncated sequence and so was dropped from phylogenetic analysis), all sequenced *amoA* genes of *Nitrospira* spp. belonged to comammox *Nitrospira* clade A (Figure S5).

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Several of the *amoA* gene sequences (i.e., RBC groups C, A, E, and D; Figure S5) clustered phylogenetically with uncultivated comammox *Nitrospira* sp. SG-bin2 from a metagenome-based survey of water treatment systems [10] and a rice paddy soil from China (Genbank accession AKD44274). The other *amoA* gene sequence groups clustered more closely to comammox *Nitrospira* from enrichment cultures or uncultivated species detected from other engineered water treatment system environments (Figure S5). Relatedness of *Nitrospira* bins based on *amoA* phylogeny (Figure S5) was similar to that obtained with the concatenated protein set, but not identical (Figure 6). For example, RBC group I (RBC069 bin) was in a clade with *Candidatus* Nitrospira nitrosa, which was also where it clustered in the phylogenomic tree. Bins for RBC group C (RBC001, RBC047) clustered separately from the *Nitrospira* from enrichment cultures, but along with *Nitrospira* sp. SG-bin2, which is not the case in the phylogenomic tree.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

This study demonstrated a dominance of comammox *Nitrospira* in biofilm samples collected from RBCs treating municipal waste of Guelph, Ontario, Canada. Not only were comammox *Nitrospira* detected as the most abundant ammonia oxidizers based on qPCR data (Figure 2), *Nitrospira* represented dominant microbial community members based on taxonomic assignments of 16S rRNA genes (Figure 3) and metagenome sequences (Figures 4, 5, 8). The comammox *Nitrospira* likely represent the majority of *Nitrospira* in the RBCs (Figures 4, 5, 8). These results contrast with previous studies of full-scale activated sludge WWTPs that have reported comammox *Nitrospira* at lower relative abundances than AOA and other AOB [7, 18–21], albeit for systems with ammonium concentrations and operational parameters distinct from those in the RBCs. This current study, along with other more recent studies [12, 15, 17, 18, 22], shows that comammox *Nitrospira* can thrive in wastewater treatment systems. Although previous analyses of the Guelph WWTP did not test for these newly discovered ammonia oxidizers [28, 29], the presence of comammox *Nitrospira* was predicted previously from activity data and an abundance of *Nitrospira* cells that actively assimilated labelled bicarbonate when incubated with ammonium [29]. The current discovery of abundant comammox *Nitrospira* communities in Guelph RBCs suggests a correspondingly important role for comammox within the 440,000 m^2 of RBC biofilm associated with this full-scale wastewater treatment system.

High relative abundances of comammox *Nitrospira* have been reported for other engineered environments with relatively low ammonium levels, such as drinking water systems. For example, comammox *Nitrospira* were found in a drinking water treatment plant [8] and distribution systems [10], and they were more abundant than other ammonia oxidizers in a groundwater well studied previously [7]. Associated with drinking water treatment plants,

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

comammox *Nitrospira* were the most abundant nitrifiers in groundwater fed rapid sand filters [13]. In those filters, clade B comammox *Nitrospira* dominated, in contrast to clade A members that we detected in the sampled RBCs. A study examining recirculating aquaculture system biofilters found that comammox *Nitrospira* were more abundant than AOA and AOB [68]. Along with these other studies, the dominance of comammox *Nitrospira* in the RBCs suggests that they compete well in engineered environments with relatively low ammonium concentrations. At the Guelph WWTP, the RBCs are located downstream of activated sludge aeration basins. Ammonium concentrations entering this tertiary treatment system are approximately two orders of magnitude lower than within aeration basins. For example, on the day of sampling in 2016, the ammonium concentration in aeration basin influents averaged 2.5 mM NH_3-N, which is typical of aerobic secondary aeration basin conditions reported elsewhere [e.g. 20–22, 69]. A large surface area for attached growth may be an important factor explaining the dominance of comammox bacteria, given their high prevalence in wastewater treatment systems with attached growth components [18]. The predicted low ammonium niche and biofilm-specific growth of comammox *Nitrospira* [4, 5] may help explain the dominance of these nitrifiers in the tertiary treatment system RBCs in Guelph.

Although relatively low ammonium concentrations likely contribute to the overall success of comammox *Nitrospira* within the RBC biofilm samples, a gradient of decreasing ammonium along the RBC flowpaths could also account for relative abundance changes of nitrifying microbial communities, both temporally and spatially. We predicted initially that comammox *Nitrospira* abundances in RBC 8 samples would exceed those in RBC 1 samples due to the predicted low ammonium niche of comammox *Nitrospira* [4], but the opposite trend was observed for both sample years (Figures 2, 5, 8). However, ammonium concentrations in the first

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RBCs were already relatively low (e.g., <18 µM for all 2016 samples; Figure 1C), and we anticipate that ammonium concentrations in RBC 8 influent (e.g., <4 µM for all 2016 samples; Figure 1C) are below the optimum for maximal predicted activity of comammox *Nitrospira*. Consistent with an "oligotrophic lifestyle", Kits *et al.* [70] reported a high apparent affinity of *Ca.* Nitrospira inopinata for ammonium ($K_{m(app)}$ of 0.65 to 1.1 µM total ammonium), and V_{max} reached by ammonium concentrations of ~5 µM, which would be similar to ammonium concentrations observed for water samples collected from RBCs sampled in this study, especially for RBC 1 stages. The affinities for ammonium of the comammox *Nitrospira* detected in the RBCs are unknown but could be consistent with the results obtained for *Ca.* N. inopinata.

Ammonium concentrations alone may not explain comammox *Nitrospira* abundances. In contrast to published studies reporting low abundances of comammox *Nitrospira* in activated sludge flocs [7, 19–21], a recent PCR-based study found that comammox *Nitrospira amoA* genes were more abundant than those of AOB in activated sludge samples from eight separate WWTPs, despite relatively high ammonium concentrations [22]. This suggests that, as seen for AOA [e.g. 71–75], the abundances of comammox *Nitrospira* vary among WWTPs. More studies are needed to identify factors that affect comammox *Nitrospira* abundances, such as substrate range, oxygen requirements, growth rates, growth yields, and biofilm formation capabilities [76].

The many unique *Nitrospira amoA* genes recovered from RBC metagenome sequences implies a high diversity of comammox *Nitrospira* within the RBC biofilm samples. Nine unique full-length comammox *Nitrospira amoA* genes were recovered and, although only three unique sequences could be placed into genome bins (Figure S5), it is likely that many of the unbinned *Nitrospira amoA* genes also represent unique comammox *Nitrospira*. Despite being unable to differentiate comammox *Nitrospira* from strict NOB in this genus using 16S rRNA genes alone

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[7], the high diversity of *Nitrospira* ASVs detected (Figure 3) is consistent with the high diversity of *Nitrospira* MAGs observed (Figure 6). Multiple populations of comammox *Nitrospira* within the same RBC system suggests that these nitrifiers do not compete directly but instead occupy unique niches within the RBCs as a result of encoded functional diversity, for example because of distinct energy sources (e.g., hydrogen, urea, cyanate) or microenvironments within the biofilm that vary in oxygen or ammonium concentrations. Indeed, the abundances of sequences that recruited to comammox *Nitrospira* (based on read recruitment to bins) varied temporally and spatially among RBCs (Figure 8), implying that complex physicochemical factors govern comammox in the RBCs. Several comammox *Nitrospira* genome MAGs may even represent closely related strains (Figure S4), adding further interest to understanding the factors influencing the distribution and ecology of these bacteria.

All detected *Nitrospira amoA* gene sequences grouped phylogenetically in comammox *Nitrospira* clade A (Figure S5), consistent with other studies reporting comammox in WWTPs [7, 12, 14–18]. Clade B comammox *Nitrospira* were not detected; the factors affecting the presence and absence of clades A and B remain unknown. Recently, Xia *et al.* [14] proposed that clade A could be further subdivided into clade A.1, which includes the cultivated comammox *Nitrospira*, and clade A.2, which contains only uncultivated comammox *amoA* sequences, including those from drinking water systems (OQW38018 and OQW37964) [10]. The *Nitrospira amoA* gene sequences detected in the RBCs were classified into both of these clade A subdivisions (Figure 6). Translated sequences of clade A.2 *amoA* genes from the RBCs were most closely related to *amoA* sequences obtained from metagenome sequences of drinking water systems [10] or to a clone library sequence from a rice paddy (Genbank accession AKD44274). Several comammox *Nitrospira* related to the rice paddy clone have been detected in other full-

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

scale WWTPs, and this clade was described as a new clade A cluster by the authors [22]. Comparisons of the dereplicated *Nitrospira* bins to reference genomes show low ANI values for many of the bins (Figure S4), which is further evidence that at least some RBC comammox bacteria and even the NOB *Nitrospira* represent novel species.

Cells of *Nitrospira* within RBC biofilm samples were previously demonstrated to assimilate labelled bicarbonate in the presence of ammonium [29], indicating autotrophic activity of these bacteria and suggesting the possibility of ammonia oxidation. In the current study, the presence of a *ure* operon (Figure 6) indicates that comammox *Nitrospira* of the RBCs can hydrolyze urea to ammonia, further supporting an ammonia-based metabolism. The presence of these genes is consistent with clade A comammox *Nitrospira* genomes and enrichment cultures that utilize urea [5, 6, 76, 77]. The comammox *Nitrospira* bins revealed that these RBC bacteria likely cannot use formate as an alternative electron donor, unlike NOB *Nitrospira* and most clade B comammox bacteria [77–80], because the comammox *Nitrospira* bins did not contain any *fdh* genes for formate dehydrogenase (Figure 6). This matches previous finding, as most clade A comammox bacteria described so far lack the genes for formate dehydrogenase [77], with the exception of *Nitrospira* bin RCA [81]. Also consistent with previous findings [18, 77], several of our comammox *Nitrospira* bins contained partial pathways for group 3b [Ni-Fe] sulfur-reducing hydrogenase genes (Figure 6), which indicates that hydrogen may serve as an alternative electron acceptor for some of these comammox *Nitrospira*.

Further supporting high functional diversity encoded RBC *Nitrospira* bins, two comammox *Nitrospira* bins, RBC069 and RBC093, contained the *cynS* gene for cyanate degradation (Figures 6, 7). Prior to this, only canonical NOB *Nitrospira* were known to possess this gene [77, 80–83]. The sublineage I *Nitrospira* RBC bins also contained the cyanase gene

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Figures 6, 7). The toxin-antitoxin gene has been postulated to indicate horizontal gene transfer in comammox *Nitrospira* [26]. The location of these genes near *cynS* in RBC093 may indicate that *cynS* was horizontally transferred to this *Nitrospira* species. The majority of cyanate containing bins are represented by *Nitrospira*, and the sublineage I *Nitrospira* make up a majority of these *cyn*-containing *Nitrospira* (Figure 7B). This indicates that *Nitrospira* represent a significant proportion of the community that are genetically capable of degrading cyanate to ammonia, either for themselves, or to supply ammonia to other nitrifiers through reciprocal feeding [82]. The *Nitrospira* in the RBCs lack traditional cyanate transporters (*cynABD*), and potentially degrade internally generated cyanate (Figure 6). This cyanate can come from the degradation of carbamoyl phosphate from the urea cycle, or from a spontaneous isomeric change of urea to cyanate [84–87]. However, it is also possible that alternate transporters could be used for uptake of cyanate, such as nitrate permease, as seen in other microorganisms [88]. This would indicate an undescribed method of nitrogen uptake for comammox bacteria.

The AOA relative abundances measured by qPCR (Figure 2) and metagenome sequence analyses (Figure 5A) differed, such that AOA were underrepresented in metagenome sequences. Previous work on the Guelph RBCs detected similar abundances of AOA in RBC samples with both qPCR and quantitative fluorescence *in situ* hybridization (FISH) [29], which suggests that the qPCR data provide a more representative estimate of AOA numbers within the sampled biofilm. Additional research would be required to assess whether low GC content of the *Ca.* N. exaquare genome (33.9%) or HMM bias against AOA sequences may contribute to underrepresentation of AOA within metagenome sequences prepared by high-throughput sequencing on the Illumina platform. Nonetheless, both qPCR and metagenomics suggested the same overall spatial and temporal distribution patterns for AOA within the RBCs.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The observed pattern of AOA *amoA* abundances for 2010 samples analyzed in this study (Figure 2, Figure 5A) was consistent with previous data [28], demonstrating again that AOA were more abundant in RBC 8 than RBC 1 samples. Although AOA in 2016 samples were at a lower relative abundance in RBC 8 compared to RBC 1, this matches data from more recent samples collected from the RBCs in December 2015 [29]. We speculate that the reversal in the AOA relative abundance pattern was caused by lower overall ammonium concentrations that have been measured in the Guelph tertiary treatment system in recent years (Figure S1B) and by cleaning of the RBC trains in 2014 that would have disturbed the microbial communities. These changes to the RBCs may have impacted the AOB as well. Both qPCR (Figure 2) and metagenome sequence data (Figure 5B) indicated an overall decrease in AOB *amoA* gene relative abundances in the 2016 samples compared to 2010. Most AOB have a relatively low ammonia affinity [70, 89, 90], and therefore the lower ammonium levels entering the RBCs in 2016 might have led to AOB being outcompeted by ammonia oxidizers with higher ammonia affinities (i.e., comammox *Nitrospira* and AOA).

This study demonstrates that comammox *Nitrospira* are consistently the dominant ammonia oxidizers in this engineered WWTP system. Detected comammox *Nitrospira* exhibited a high level of overall diversity, and much higher diversity than the AOA in the same systems. The evolutionary history of comammox *Nitrospira* prior to establishing within the WWTP, perhaps within heterogeneous soil or sediment samples, may be a factor that contributed to the high diversity observed within RBCs. There is some phylogenetic evidence that the AOA selected within the RBCs may have originated on skin surfaces [29, 91], which would offer relatively low habitat heterogeneity and could help explain the low diversity observed in the RBC system. Although comammox *Nitrospira* were numerically dominant, their contributions to

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

nitrification in the RBC environment remain unclear. Future studies using differential inhibitors, isotope tracer studies, and laboratory cultivation may be useful for elucidating the contributions of comammox *Nitrospira* to nitrification in the RBC biofilms. In addition, the comammox *Nitrospira* that fall within the newly proposed clade A.2 [14] would be ideal targets for cultivation given that no cultured representatives are yet available for this group.

The Guelph WWTP RBCs represent a unique and useful system to study the ecology of comammox *Nitrospira* as a result of built-in environmental gradients and the combined presence of comammox *Nitrospira*, AOA, AOB, and NOB. Tertiary treatment systems such as these RBCs can be used to produce high quality effluents, which may be particularly important when WWTPs discharge into environments that have low assimilative capacities or are ecologically sensitive. Understanding the microorganisms that are present and their contributions to nitrification within the context of water treatment is an important step towards understanding the microbial contributions to nitrification in low-ammonium biofilm systems and may contribute to knowledge of operational practices for improved effluent quality in municipal, industrial, and aquaculture-associated water treatment systems.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We thank the operators and staff at the Guelph WWTP for access to their facility and for providing plant data. We thank the Joint Genome Institute for support and performance of the amplicon sequencing and The Centre for Applied Genomics for performing the metagenomic sequencing. Funding for metagenome sequencing was provided by the Faculties of Science and Engineering at the University of Waterloo. This research was supported by Discovery Grants to LAH, ACD, WJP, and JDN from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Competing interests

The authors declare no competing interests.

References

1. Daims H, Lücker S, Wagner M. A new perspective on microbes formerly known as nitrite-oxidizing bacteria. *Trends Microbiol* 2016; **24**: 699–712.

2. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 2005; **437**: 543–546.

3. Purkhold U, Pommerening-Röser A, Juretschko S, Schmid MC, Koops HP, Wagner M. Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: Implications for molecular diversity surveys. *Appl*

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Environ Microbiol 2000; **66**: 5368–5382.

4. Costa E, Pérez J, Kreft J-U. Why is metabolic labour divided in nitrification? *Trends Microbiol* 2006; **14**: 213–219.

5. Daims H, Lebedeva E V., Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by *Nitrospira* bacteria. *Nature* 2015; **528**: 504–509.

6. van Kessel MAHJ, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, et al. Complete nitrification by a single microorganism. *Nature* 2015; **528**: 555–559.

7. Pjevac P, Schauberger C, Poghosyan L, Herbold CW, Kessel MA van, Daebeler A, et al. *AmoA*-targeted polymerase chain reaction primers for the specific detection and quantification of comammox *Nitrospira* in the environment. *Front Microbiol* 2017; **8**: 1508.

8. Pinto AJ, Marcus DN, Ijaz Z, Bautista-de los Santos QM, Dick GJ, Raskin L. Metagenomic evidence for the presence of comammox *Nitrospira*-like bacteria in a drinking water system. *mSphere* 2015; **1**: e00054-15.

9. Palomo A, Fowler SJ, Gülay A, Rasmussen S, Sicheritz-Ponten T, Smets BF. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of *Nitrospira* spp. *ISME J* 2016; **10**: 2569–2581.

10. Wang Y, Ma L, Mao Y, Jiang X, Xia Y, Yu K, et al. Comammox in drinking water systems. *Water Res* 2017; **116**: 332–341.

11. Palomo A, Dechesne A, Smets BF. Genomic profiling of *Nitrospira* species reveals ecological success of comammox *Nitrospira*. *bioRxiv* 2019; 612226.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Zhao Z, Huang G, Wang M, Zhou N, He S, Dang C, et al. Abundance and community composition of comammox bacteria in different ecosystems by a universal primer set. *Sci Total Environ* 2019; **691**: 146–155.

13. Fowler SJ, Palomo A, Dechesne A, Mines PD, Smets BF. Comammox *Nitrospira* are abundant ammonia oxidizers in diverse groundwater-fed rapid sand filter communities. *Environ Microbiol* 2018; **20**: 1002–1015.

14. Xia F, Wang J-G, Zhu T, Zou B, Rhee S-K, Quan Z-X. Ubiquity and diversity of complete ammonia oxidizers (comammox). *Appl Environ Microbiol* 2018; **84**: e01390-18.

15. Beach NK, Noguera DR. Design and assessment of species-level qPCR primers targeting comammox. *Front Microbiol* 2019; **10**: 36.

16. Annavajhala MK, Kapoor V, Santo-Domingo J, Chandran K. Comammox functionality identified in diverse engineered biological wastewater treatment systems. *Environ Sci Technol Lett* 2018; **5**: 110–116.

17. Roots P, Wang Y, Rosenthal AF, Griffin JS, Sabba F, Petrovich M, et al. Comammox *Nitrospira* are the dominant ammonia oxidizers in a mainstream low dissolved oxygen nitrification reactor. *Water Res* 2019; **157**: 396–405.

18. Cotto I, Dai Z, Huo L, Anderson CL, Vilardi KJ, Ijaz U, et al. Long solids retention times and attached growth phase favor prevalence of comammox bacteria in nitrogen removal systems. *bioRxiv* 2019; 696351.

19. Chao Y, Mao Y, Yu K, Zhang T. Novel nitrifiers and comammox in a full-scale hybrid biofilm and activated sludge reactor revealed by metagenomic approach. *Appl Microbiol*

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Biotechnol 2016; **100**: 8225–8237.

20. Fan X-Y, Gao J-F, Pan K-L, Li D-C, Dai H-H. Temporal dynamics of bacterial communities and predicted nitrogen metabolism genes in a full-scale wastewater treatment plant. *RSC Adv* 2017; **7**: 56317–56327.

21. Pan KL, Gao JF, Fan XY, Li DC, Dai HH. The more important role of archaea than bacteria in nitrification of wastewater treatment plants in cold season despite their numerical relationships. *Water Res* 2018; **145**: 552–561.

22. Wang M, Huang G, Zhao Z, Dang C, Liu W, Zheng M. Newly designed primer pair revealed dominant and diverse comammox *amoA* gene in full-scale wastewater treatment plants. *Bioresour Technol* 2018; **270**: 580–587.

23. Zheng M, Wang M, Zhao Z, Zhou N, He S, Liu S, et al. Transcriptional activity and diversity of comammox bacteria as a previously overlooked ammonia oxidizing prokaryote in full-scale wastewater treatment plants. *Sci Total Environ* 2019; **656**: 717–722.

24. Gonzalez-Martinez A, Rodriguez-Sanchez A, van Loosdrecht MCM, Gonzalez-Lopez J, Vahala R. Detection of comammox bacteria in full-scale wastewater treatment bioreactors using tag-454-pyrosequencing. *Environ Sci Pollut Res* 2016; **23**: 25501–25511.

25. Fan X-Y, Gao J-F, Pan K-L, Li D-C, Zhang L-F, Wang S-J. Shifts in bacterial community composition and abundance of nitrifiers during aerobic granulation in two nitrifying sequencing batch reactors. *Bioresour Technol* 2018; **251**: 99–107.

26. Camejo PY, Santo Domingo J, McMahon KD, Noguera DR. Genome-enabled insights

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

into the ecophysiology of the comammox bacterium "*Candidatus* Nitrospira nitrosa". *mSystems* 2017; **2**: e00059-17.

27. Metch JW, Wang H, Ma Y, Miller JH, Vikesland PJ, Bott C, et al. Insights gained into activated sludge nitrification through structural and functional profiling of microbial community response to starvation stress. *Environ Sci Water Res Technol* 2019; **5**: 884–896.

28. Sauder LA, Peterse F, Schouten S, Neufeld JD. Low-ammonia niche of ammonia-oxidizing archaea in rotating biological contactors of a municipal wastewater treatment plant. *Environ Microbiol* 2012; **14**: 2589–2600.

29. Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of *Candidatus* Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* 2017; **11**: 1142–1157.

30. Holmes RM, Aminot A, Kérouel R, Hooker BA, Peterson BJ. A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can J Fish Aquat Sci* 1999; **56**: 1801–1808.

31. Poulin P, Pelletier É. Determination of ammonium using a microplate-based fluorometric technique. *Talanta* 2007; **71**: 1500–1506.

32. Miranda KM, Espey MG, Wink DA. A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric Oxide* 2001; **5**: 62–71.

33. Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

time PCR. *Environ Microbiol* 2003; **5**: 787–797.

34. Muyzer G, de Waal EC, Uitterlinden AG. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 1993; **59**: 695–700.

35. Rotthauwe JH, Witzel KP, Liesack W. The ammonia monooxygenase structural gene *amoA* as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl Environ Microbiol* 1997; **63**: 4704–4712.

36. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016; **18**: 1403–1414.

37. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 2011; **12**: 38.

38. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Prepr* 2018; **6**: e27295v2.

39. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; **17**: 10.

40. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.

41. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2012; **41**: D590–D596.

42. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011; **12**: 2825–2830.

43. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 2016; **9**: 88.

44. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. .

45. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047–3048.

46. Kim D, Hahn AS, Wu SJ, Hanson NW, Konwar KM, Hallam SJ. FragGeneScan-plus for scalable high-throughput short-read open reading frame prediction. *2015 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2015*. 2015. IEEE, pp 1–8.

47. Fish JA, Chai B, Wang Q, Sun Y, Brown CT, Tiedje JM, et al. FunGene: the functional gene pipeline and repository. *Front Microbiol* 2013; **4**: 291.

48. Petrenko P, Lobb B, Kurtz DA, Neufeld JD, Doxey AC. MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. *BMC Biol* 2015; **13**: 92.

49. White RAI, Brown J, Colby S, Overall CC, Lee J-Y, Zucker J, et al. ATLAS (Automatic Tool for Local Assembly Structures) -a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. *PeerJ Prepr* 2017; **5**: e2843v1.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

50. Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *bioRxiv* 2019; 737528.

51. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018; **36**: 996–1004.

52. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; **9**: 5114.

53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**: 421.

54. Bergstrand LH, Cardenas E, Holert J, van Hamme JD, Mohn WW. Delineation of steroid-degrading microorganisms through comparative genomic analysis. *MBio* 2016; **7**: e00166.

55. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019.

56. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32**: 268–274.

57. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14**: 587–589.

58. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol* 2011; **7**: e1002195.

59. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

60. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; **33**: 1870–1874.

61. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008; **25**: 1307–1320.

62. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high□quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011; **7**: 539.

63. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 2018; **27**: 135–145.

64. Kamennaya NA, Post AF. Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* 2011; **77**: 291–301.

65. Pester M, Maixner F, Berry D, Rattei T, Koch H, Lücker S, et al. *NxrB* encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing *Nitrospira*. *Environ Microbiol* 2014; **16**: 3055–3071.

66. Dahllof I, Baillie H, Kjelleberg S. *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol* 2000; **66**: 3376–3380.

67. Walsh MA, Otwinowski Z, Perrakis A, Anderson PM, Joachimiak A. Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. *Structure* 2000; **8**: 505–14.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

68. Bartleme RP, Mclellan SL, Newton RJ. 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**: 101.

69. Park H-D, Wells GF, Bae H, Criddle CS, Francis CA. Occurrence of ammonia-oxidizing archaea in wastewater treatment plant bioreactors. *Appl Environ Microbiol* 2006; **72**: 5643–5647.

70. Kits KD, Sedlacek CJ, Lebedeva E V., Han P, Bulaev A, Pjevac P, et al. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* 2017; **549**: 269–272.

71. Bai Y, Sun Q, Wen D, Tang X. Abundance of ammonia-oxidizing bacteria and archaea in industrial and domestic wastewater treatment systems. *FEMS Microbiol Ecol* 2012; **80**: 323–330.

72. Roy D, McEvoy J, Blonigen M, Amundson M, Khan E. Seasonal variation and ex-situ nitrification activity of ammonia oxidizing archaea in biofilm based wastewater treatment processes. *Bioresour Technol* 2017; **244**: 850–859.

73. Gao J-F, Luo X, Wu G-X, Li T, Peng Y-Z. Quantitative analyses of the composition and abundance of ammonia-oxidizing archaea and ammonia-oxidizing bacteria in eight full-scale biological wastewater treatment plants. *Bioresour Technol* 2013; **138**: 285–296.

74. Zhang T, Ye L, Tong AHY, Shao MF, Lok S. Ammonia-oxidizing archaea and ammonia-oxidizing bacteria in six full-scale wastewater treatment bioreactors. *Appl Microbiol Biotechnol* 2011; **91**: 1215–1225.

75. Limpiyakorn T, Sonthiphand P, Rongsayamanont C, Polprasert C. Abundance of *amoA*

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genes of ammonia-oxidizing archaea and bacteria in activated sludge of full-scale wastewater treatment plants. *Bioresour Technol* 2011; **102**: 3694–3701.

76. Lawson CE, Lücker S. Complete ammonia oxidation: an important control on nitrification in engineered ecosystems? *Curr Opin Biotechnol* 2018; **50**: 158–165.

77. Palomo A, Pedersen AG, Fowler SJ, Dechesne A, Sicheritz-Pontén T, Smets BF. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox *Nitrospira*. *ISME J* 2018; **12**: 1779–1793.

78. Koch H, Lücker S, Albertsen M, Kitzinger K, Herbold C, Spieck E, et al. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc Natl Acad Sci U S A* 2015; **112**: 11371–11376.

79. Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, 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**: 13479–13484.

80. Ushiki N, Fujitani H, Shimada Y, Morohoshi T, Sekiguchi Y, Tsuneda S. Genomic analysis of two phylogenetically distinct *Nitrospira* species reveals their genomic plasticity and functional diversity. *Front Microbiol* 2018; **8**: 2637.

81. Poghosyan L, Koch H, Lavy A, Frank J, Kessel MAHJ, Jetten MSM, et al. Metagenomic recovery of two distinct comammox *Nitrospira* from the terrestrial subsurface. *Environ Microbiol* 2019; E-pub DOI 10.1111/1462-2920.14691.

82. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate as an energy source for nitrifiers. *Nature* 2015; **524**: 105–108.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

83. Koch H, van Kessel MAHJ, Lücker S. Complete nitrification: insights into the ecophysiology of comammox *Nitrospira*. *Appl Microbiol Biotechnol* 2019; **103**: 177–189.

84. Marier JR, Rose D. Determination of cyanate, and a study of its accumulation in aqueous solutions of urea. *Anal Biochem* 1964; **7**: 304–314.

85. Dirnhuber P, Schütz F. The isomeric transformation of urea into ammonium cyanate in aqueous solutions. *Biochem J* 1948; **42**: 628–32.

86. Anderson PM, Sung Y, Fuchs JA. The cyanase operon and cyanate metabolism. *FEMS Microbiol Lett* 1990; **87**: 247–252.

87. Purcarea C, Ahuja A, Lu T, Kovari L, Guy HI, Evans DR. *Aquifex aeolicus* aspartate transcarbamoylase, an enzyme specialized for the efficient utilization of unstable carbamoyl phosphate at elevated temperature. *J Biol Chem* 2003; **278**: 52924–34.

88. Muñoz-Centeno MC, Paneque A, Cejudo FJ. Cyanate is transported by the nitrate permease in *Azotobacter chroococcum*. *FEMS Microbiol Lett* 1996; **137**: 91–94.

89. Lehtovirta-Morley LE. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiol Lett* 2018; **365**: fny058.

90. Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: The quest for niche specialisation and differentiation. *Trends Microbiol* 2012; **20**: 523–531.

91. Probst AJ, Auerbach AK, Moissl-Eichinger C. Archaea on human skin. *PLoS One* 2013; **8**: e65388.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Figure 1 Rotating biological contactors (RBCs) of the Guelph wastewater treatment plant (WWTP). (A) External view of the RBCs and (B) schematic of the RBC trains. Arrows indicate the direction of water flow. (C) Ammonium (NH_4^+-N), nitrite (NO_2-N), and nitrate (NO_3-N) concentrations in RBC influent water sampled at the same time as biofilm collection in October 2016. Error bars indicate standard deviation of technical duplicates. NE, Northeast train; NW, Northwest train; SE, Southeast train; SW, Southwest train.

Figure 2 Bacterial 16S rRNA, comammox *Nitrospira* (*amoA*), AOB (*amoA*), and thaumarcheotal 16S rRNA gene abundances for samples paired with metagenome sequencing. Error bars indicate standard deviation of technical qPCR duplicates. Pie charts show gene abundances as a proportion of all ammonia oxidizing prokaryotes. Numbers above the pie charts indicate the proportion (%) of the total community that all ammonia oxidizing prokaryotes represent. Gene copies were calculated based on the amount of DNA present in the original extractions but were not standardized to account for expected gene copy number per genome. NE, northeast train; NW, northwest train; SE, southeast train; SW, southwest train.

Figure 3 Relative abundances of amplicon sequence variants (ASVs) based on 16S rRNA gene amplicon sequencing, for samples paired with metagenome sequencing. Only taxa detected at ≥ 1% relative abundance are shown. The lowest level of informative taxonomic rank is given. Numbers after the underscore are the ASV number. The numbers inside the circles represent the relative abundance, in percent. *Nitrospira* ASVs are highlighted in orange, AOB ASVs are highlighted in green, and the AOA ASV is highlighted in blue.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4 Taxonomic profiling of RBC microbial community by the hidden Markov model (HMM) for the RNA polymerase beta subunit (*rpoB*). Stacked bars represent the relative abundance of unassembled *rpoB* metagenome sequence reads classified at the genus level. Genera at greater than or equal to 1% relative abundance are shown. Dark orange and green shaded bars represent nitrifiers. *Nitrospira* species include both strict nitrite-oxidizing and comammox *Nitrospira*, as they cannot be distinguished at the genus level. NE, northeast train; NW, northwest train; SE, southeast train; SW, southwest train.

Figure 5 Functional gene profiling of nitrifiers in rotating biological contactor (RBC) microbial communities by hidden Markov models (HMMs). Stacked bars represent the normalized abundances of functional genes, relative to *rpoB*, from unassembled metagenome sequence reads. Genera at ≥1% normalized relative abundance are shown with taxa labels, along with genera at <1%, which are marked as "Other" at the uppermost grey bar. Blue, green, and orange bars represent nitrifiers. Other genera at ≥1% normalized relative abundance are collapsed to the phylum level for clarity and are shown in grey. (A) *amoA* gene of ammonia-oxidizing archaea, which are grouped at the phylum level (*Thaumarchaeota*) for clarity. No HMM hits were found for sample NE1 June 2010. (B) *amoA* gene of ammonia-oxidizing bacteria and comammox *Nitrospira*. (C) *nxrB* gene of *Nitrospira* and other NOB. NE, northeast train; NW, northwest train; SE, southeast train; SW, southwest train; *amoA*, ammonia monooxygenase subunit A gene; *nxrB*, nitrite oxidoreductase.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6 Phylogeny and gene pathways in *Nitrospira* bins and reference genomes. The bins recovered from this study are indicated in bold, an asterisk indicates that the genome is from an enrichment or pure culture. Phylogenomic analysis was performed using a concatenated set of 74 core bacterial proteins. All bootstrap values are 100% except those shown, and the scale bar represents the proportion of amino acid change. Gene annotation was determined using reciprocal BLASTP against three reference *Nitrospira* genomes (*Ca.* N. inopinata, *Nitrospira moscoviensis*, *Nitrospira lenta*), and amino acid identity to the reference gene is indicated in the heatmap. Abbreviations: *amo*, ammonia monoxygenase; *Rh50*, ammonia transporter; *amt*, ammonia transporter; *hao*, hydroxylamine oxidoreductase; *cyc*, cytochrome c; *nxr*, nitrite oxidoreductase; *ure*, urease; *urt*, urea trasporter; *nrf*, cytochrome c nitrite reductase; *nirK*, dissimilatory nitrite reductase; *nirA*, assimilatory nirite reductase; *narK*, nitrute/nitrate transporter; *nirC*, nitrite transporter; *cynABD*, cyanate transporter; *cynS*, cyanate hydratase; *hup*, group 2a [Ni-Fe] hydrogenase; *hyp*, hydrogenase accessory protein; *hyf*, putative group 4 hydrogenase; *hyb* and *hyd*, group 3 [Ni-Fe] sulfur-reducing hydrogenase; *fdh*, formate dehydrogenase.

Figure 7 Phylogeny and relative abundances of cyanases (CynS) detected in RBC metagenomic data. (A) Maximum likelihood phylogeny of CynS primary sequences detected in *Nitrospira* genome bins from this study compared to reference sequences from NCBI. Sublineages of the *Nitrospira* clade are highlighted by the coloured vertical bar, including potential clade A comammox bacteria, which are highlighted in orange. Genome bins recovered from the RBCs are bolded, and an asterisk appears after the names of genomes of cultivated organisms. Bootstrap values over 50% are shown, and the scale bar represents the proportion of amino acid

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

change. To the right of the phylogeny, a portion of the CynS primary sequence alignment is shown along with overall sequence similarity based on the BLOSUM62 metric. Residues with 100% conservation are marked with an asterisk. Blue asterisks are used to mark known catalytic residues of CynS, namely Arg96, Glu99, and Ser122. (B) Relative abundances of all RBC genome bins containing *cynS* among RBC metagenomes. Genome bins are collapsed at the genus level into stacked bars, and relative abundances are based on the propotion (%) of mapped assembled reads to genome data.

Figure 8 Relative abundance of *Nitrospira* bins across samples based on total recruited reads. The size of the bubbles indicates the proportion (%) of assembled reads from each sample that mapped to each genome bin. Stacked bars summarize the total proportions of assembled metagenomic reads recruited to the genome bins.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1 Bin statistics and classification of *Nitrospira* metagenome assembled genomes (MAGs).

MAG ID	Completeness (%)	Contamination (%)	Genome size (Mbp)	Number of contigs	Contig N50	GC content (%)	*Nitrospira* lineage
RBC001	94.9	2.7	3.8	227	32534	54.9	Clade A comammox
RBC003	94.0	1.8	4.3	295	24838	50.1	Sublineage IV
RBC021	78.5	1.3	3.6	460	11365	54.6	Clade A comammox
RBC026	75.1	3.2	2.9	567	7228	58.9	Sublineage I
RBC035	94.9	2.3	4.5	473	14062	55.0	Clade A comammox
RBC042	89.6	4.8	3.4	469	11565	54.7	Clade A comammox
RBC044	89.9	0.9	3.6	343	15917	54.7	Clade A comammox
RBC047	75.7	3.5	3.0	542	7102	54.5	Clade A comammox
RBC048	94.9	2.7	3.9	174	38558	58.6	Sublineage I
RBC069	95.9	2.7	4.5	87	78382	54.5	Clade A comammox
RBC073	91.3	0.0	4.0	64	161412	59.5	Sublineage I
RBC083	82.4	1.9	2.7	212	21077	55.0	Clade A comammox
RBC085	92.1	1.8	3.7	250	25628	59.0	Sublineage I
RBC093	94.9	2.8	3.8	58	116502	55.0	Clade A comammox
RBC100	94.0	2.3	4.3	233	32473	55.2	Clade A comammox

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

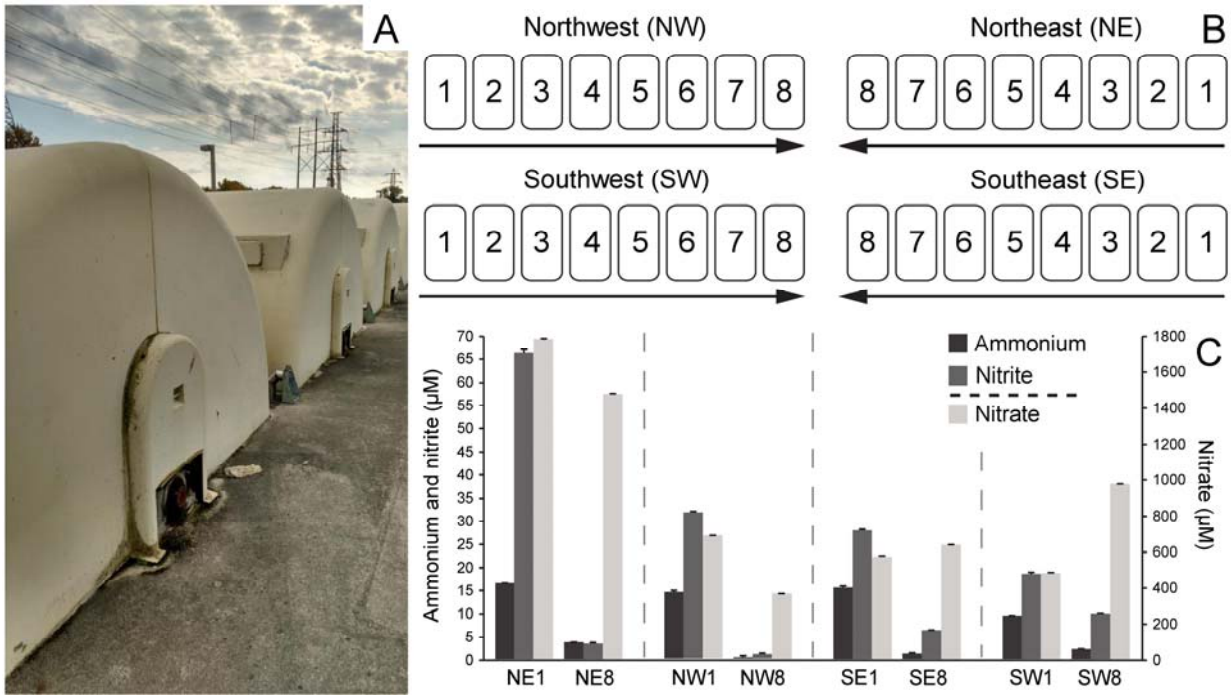


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

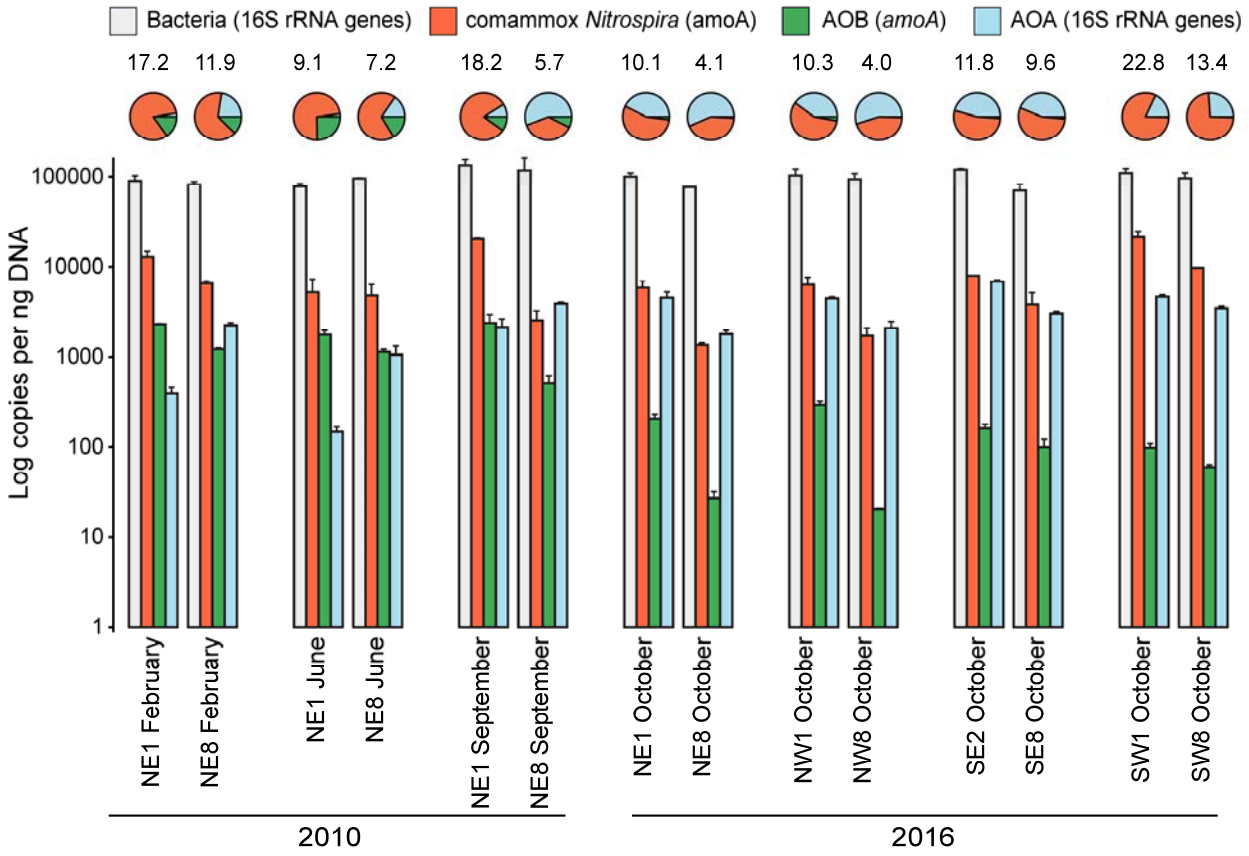


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1020

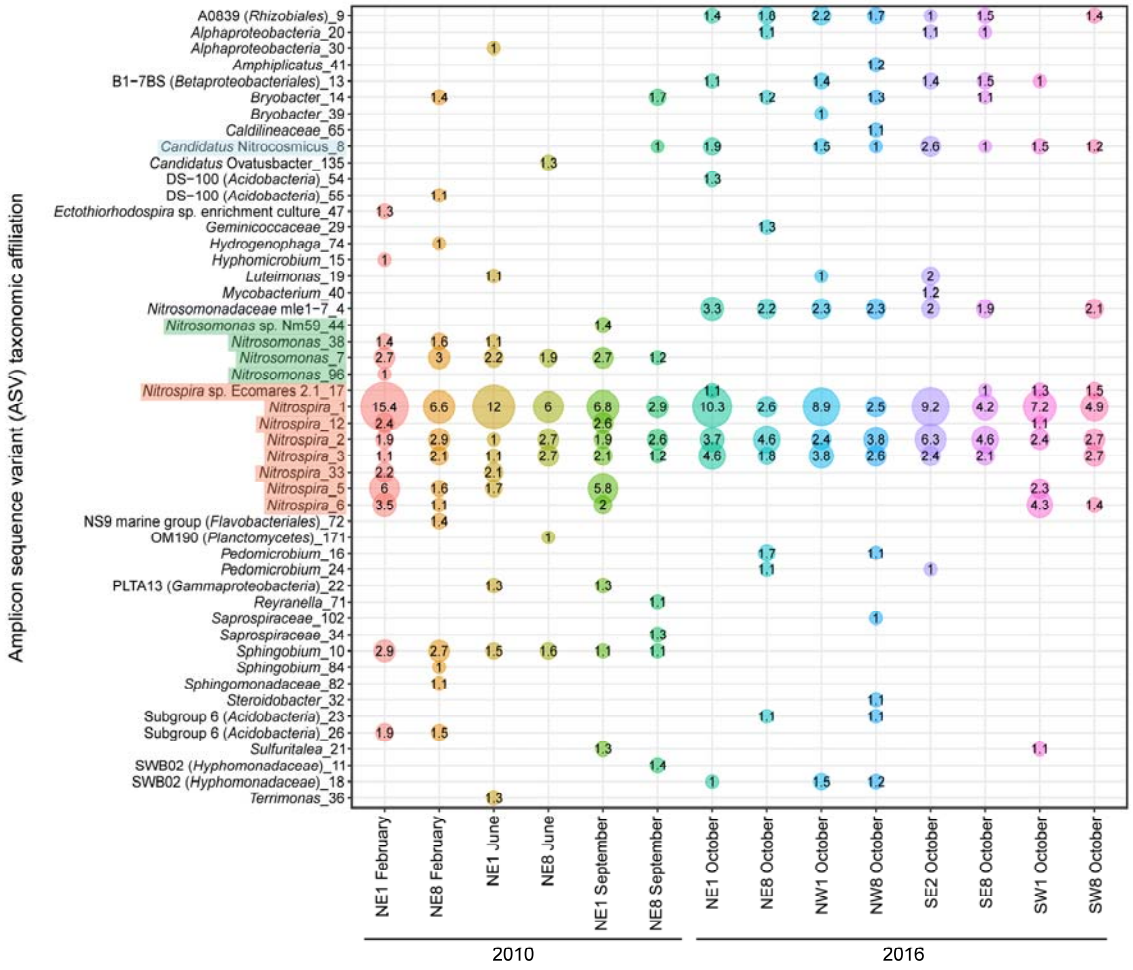
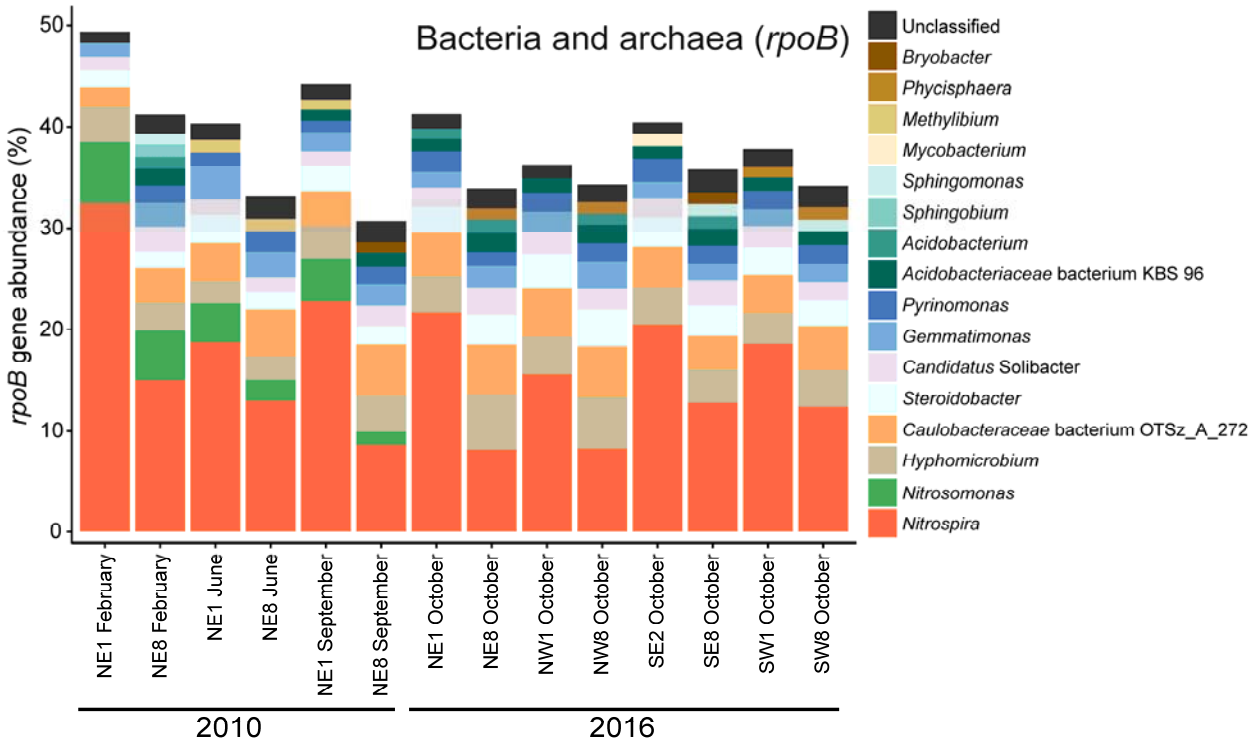


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1025



1030

Figure 4

1035

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1040

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

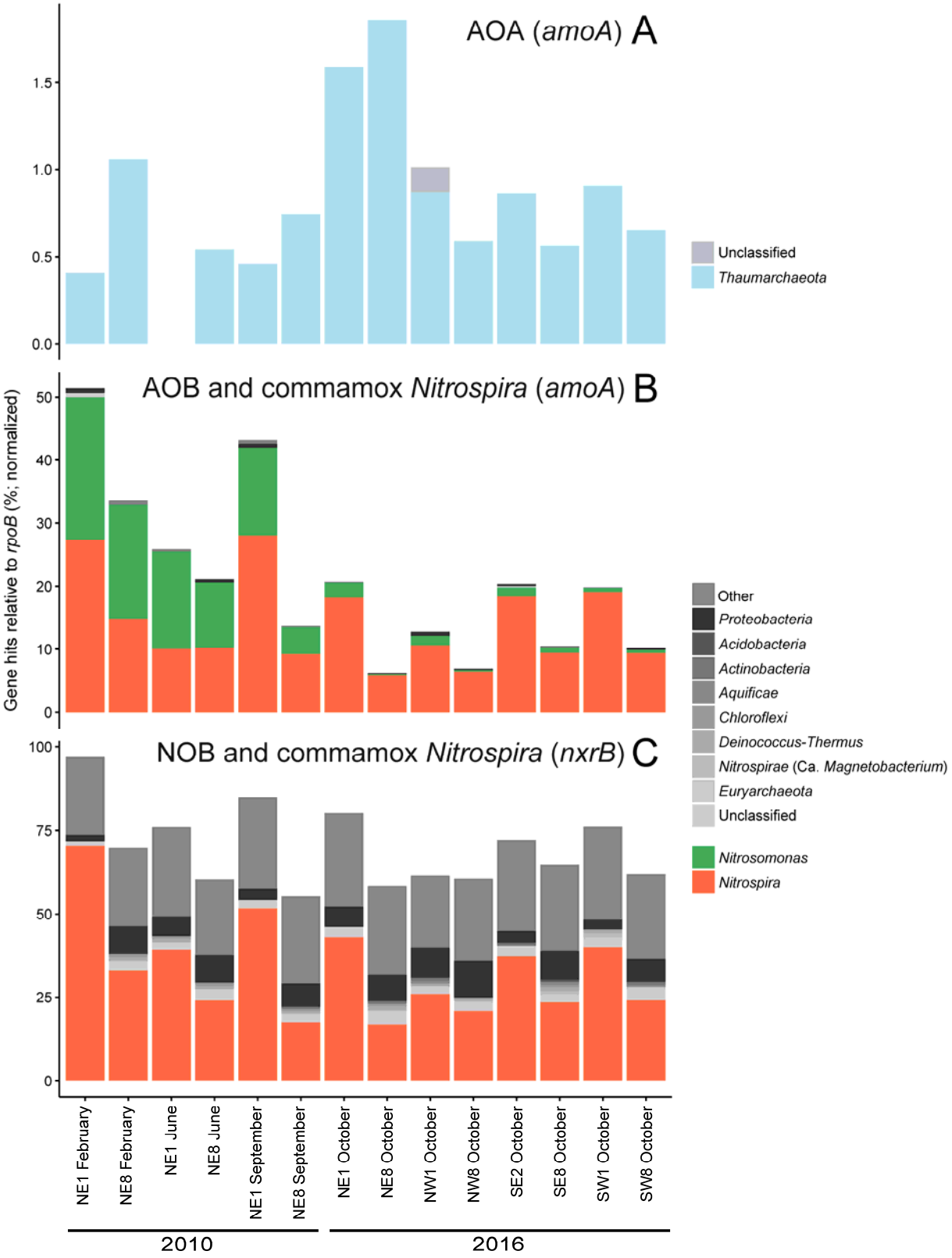
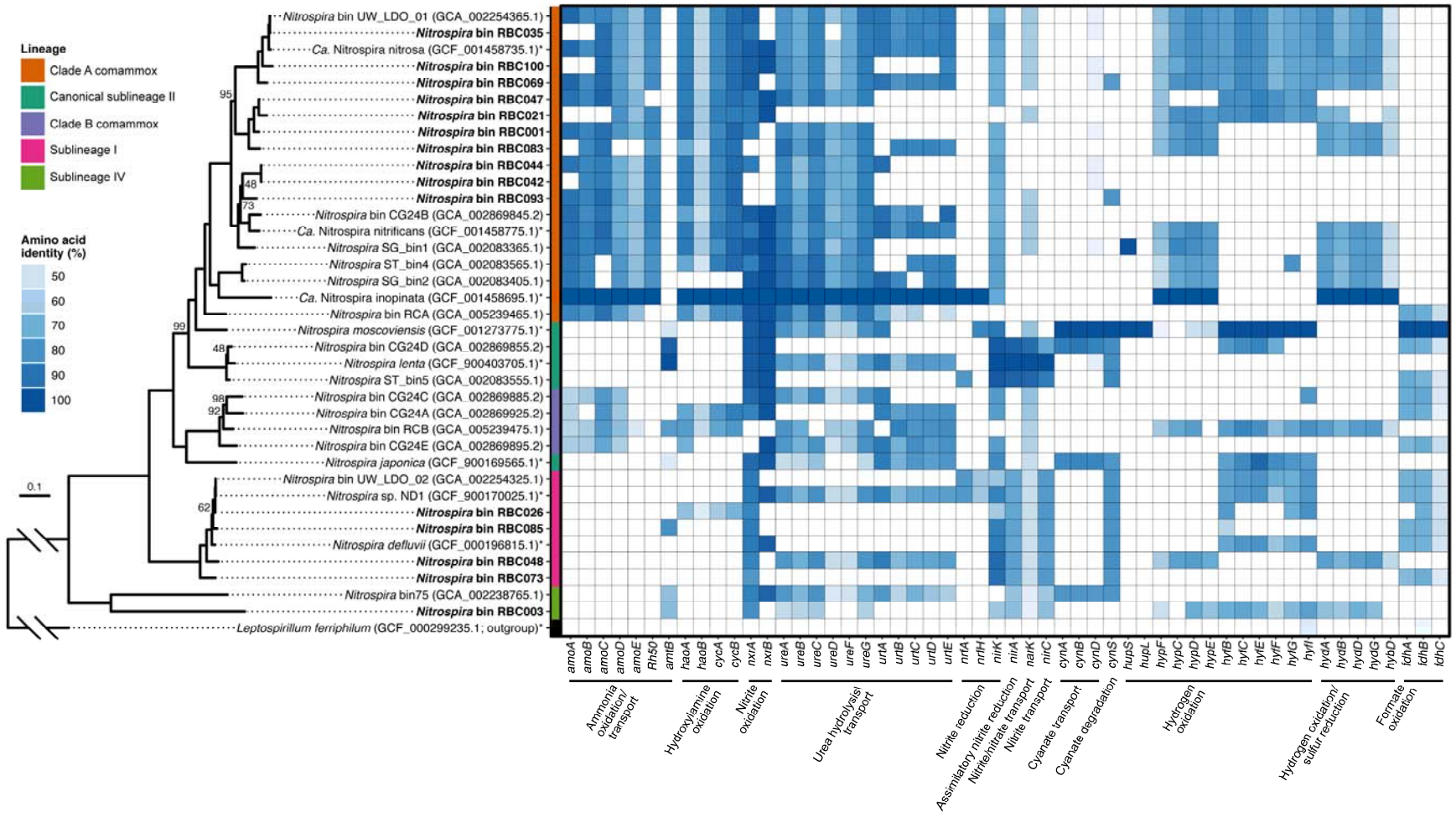


Figure 5



1045 **Figure 6**

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

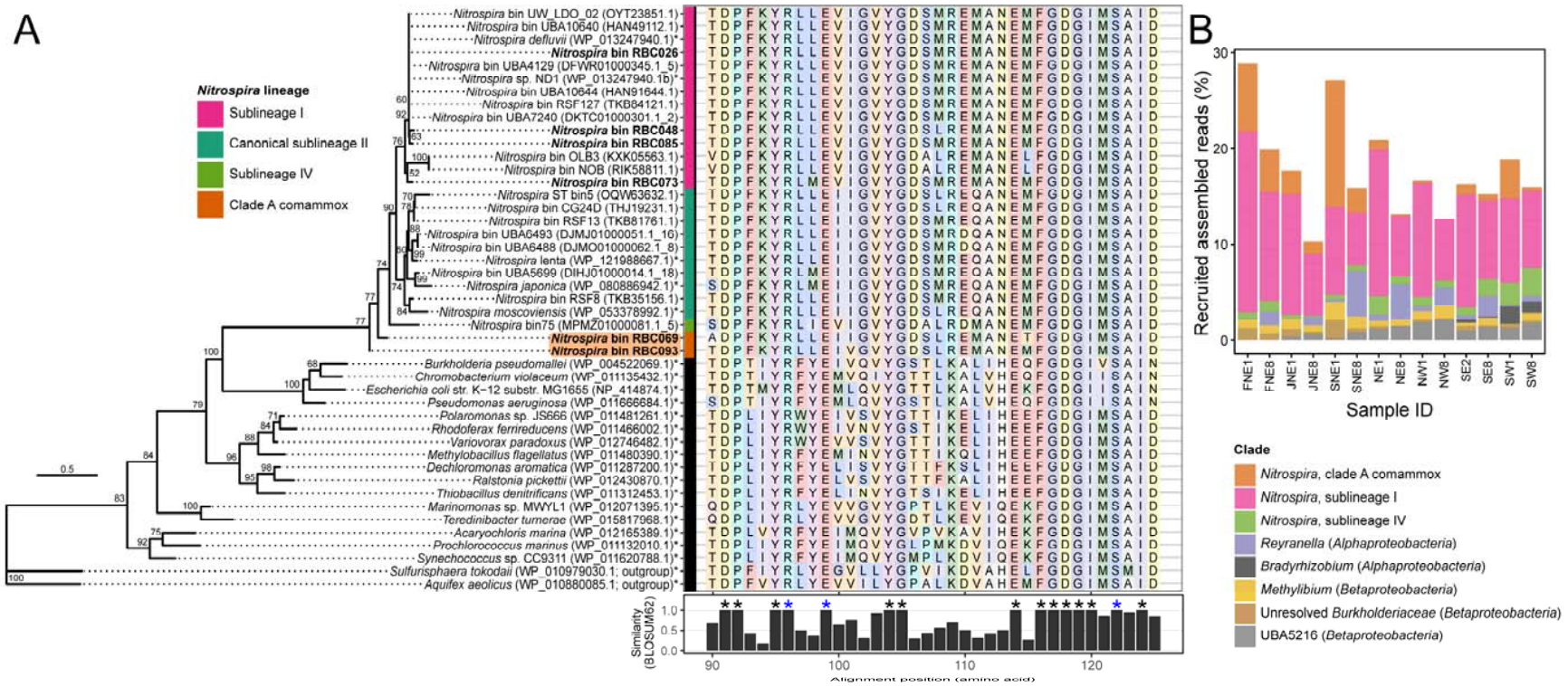
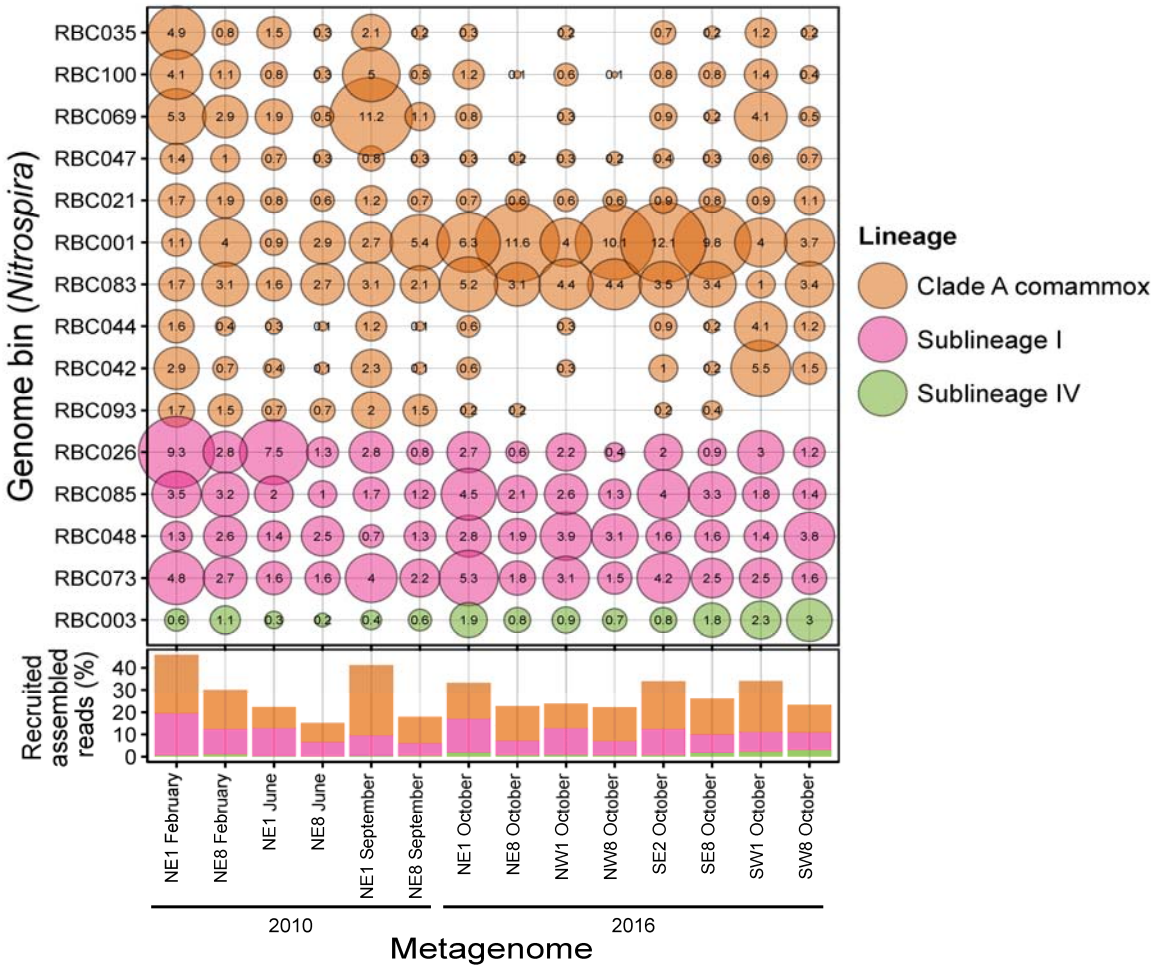


Figure 7

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/529826; this version posted September 27, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



1050

Figure 8