High functional diversity among *Nitrospira* populations that dominate rotating biological contactor microbial 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],

5 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

10 *These authors contributed equally

#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

15 Running title: Comammox *Nitrospira* dominate wastewater biofilm

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

1

Abstract

20   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-

25   oxidizing bacteria (AOB) in biofilm samples collected from tertiary 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 dominate RBC biofilm samples and that comammox *Nitrospira* outnumber all other nitrifiers. Genome bins recovered from assembled metagenomes reveal multiple

30   populations of comammox *Nitrospira* with distinct spatial and temporal distributions, including several taxa that are distinct from previously characterized *Nitrospira* members. 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. Our metagenome bins also reveal two cyanase-encoding

35   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 as model systems for continued investigation of environmental factors that control the distributions and activities of AOB, AOA, comammox *Nitrospira*, and other nitrite oxidizers.

40 **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 sequential oxidation of ammonia to nitrate, via nitrite, and these two enzymatic steps were historically thought to be mediated by distinct

45 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]. Microorganisms capable of oxidizing both ammonia and nitrite via complete ammonia oxidation (comammox) were predicted over a decade ago to be slow growing and to inhabit biofilms exposed to relatively low ammonium concentrations [4]. These

50 predictions were confirmed by the discovery of *Nitrospira* members capable of catalyzing comammox [5, 6], and all known comammox *Nitrospira* 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].

55 Compared to AOA and AOB, 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

60 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 in some WWTPs [12, 18, 22], and comammox *Nitrospira* have been enriched in wastewater

3

treatment reactors with low dissolved oxygen conditions [15, 17]. Additionally, the high abundance of *amoA* transcripts from comammox *Nitrospira* detected in activated sludge suggests

65 an active role in 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 our knowledge, no studies have investigated comammox *Nitrospira* in biofilm-based tertiary wastewater treatment systems.

The current study examined full scale tertiary rotating biological contactors (RBCs;

70 Figure 1A) 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, the RBCs are organized into four "trains" that are each composed of eight individual RBC "stages" (Figure 1B). The RBC-associated AOA and AOB communities were characterized previously, demonstrating that the sole detectable AOA population increased in abundance along

75 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 of *Nitrospira* cells, as viewed by fluorescence *in situ* hybridization (FISH), raised the possibility that some of the detected *Nitrospira* may contribute to comammox [29]. Due to the predicted low ammonium niche and biofilm growth of

80 comammox bacteria [4, 5], we hypothesized that comammox *Nitrospira* would dominate the RBC biofilm and that the relative abundance of comammox *Nitrospira* would increase as ammonium concentrations decrease 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

4

85   tertiary treatment system, using a combination of quantitative PCR (qPCR), 16S rRNA gene

sequencing, and metagenomics.

Materials and Methods

Sampling

Using ethanol-cleaned spatulas, biofilm was sampled from the RBC trains in Guelph,

90   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 biofilm

samples that were collected previously for temporal comparison [28]. These reference samples

95   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, kept on ice, and

then frozen at -20°C until chemical analysis.

Water chemistry

Ammonium was measured fluorometrically using orthophthaldialdehyde (OPA) reagent

100   [30] as described previously [31] and nitrite/nitrate was assessed colorimetrically with the Greiss

reagent [32]. Measurements represent total ammonium as nitrogen, and nitrite and nitrate as

nitrogen; details are described in the supplemental methods. 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

105   compiled for the years 2010 to 2017. Plant influent and effluent data were retrieved from Guelph

WWTP annual reports for 2010 to 2017

(https://guelph.ca/living/environment/water/wastewater/).

DNA extractions

Extractions were performed from 0.25 g (wet weight) of biofilm with the PowerSoil

110   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 the NanoDrop 2000 (Thermo

Scientific, Waltham, MA, USA) and Qubit dsDNA HS 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

115   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

120   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). 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),

125   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

6

primers produced no amplicons (data not shown). All qPCR amplifications were carried out as technical duplicates on a CFX96 Real-Time PCR Detection System (Bio-Rad, Hercules, CA,

130   USA). Additional thermal cycling conditions and standards are described in the supplemental methods.

16S rRNA gene amplicon sequencing

The V4-V5 regions of 16S rRNA genes were amplified using primers 515F [36] and 926R [37] with Illumina adapters. The sequence data were produced by the US Department of

135   Energy Joint Genome Institute (JGI) using standard operating procedures. Triplicates were combined, quantified, and sample amplicons were then pooled equally. The pooled library was sequenced on a MiSeq (Illumina) with 2x300 base reads. Sequence analysis was performed with QIIME2 version 2019.1.0 [38]. Sequences were trimmed to remove primer and adapter sequences with cutadapt [39]. Quality trimming, denoising, error-correction, paired-end read

140   merging, chimera removal, and dereplication was performed with DADA2 [40], producing an amplicon sequence variant (ASV) table with 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].

Metagenome sequencing and analysis

145   Shearing of DNA, library preparation, and sequencing were performed at The Centre for Applied Genomics (TCAG) in Toronto, Ontario. Extracted DNA was quantified using the Qubit dsDNA HS Assay kit (Thermo Scientific, Waltham, MA, USA), and 500 ng of input DNA was sheared into ~550 bp fragments using a LE220 Focused-ultrasonicator (Covaris, Woburn, MA, USA). Library preparation was performed using the TruSeq PCR-free Library Prep Kit

7

150   (Illumina). 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 were performed using

AdapterRemoval version 2.2.2 [43] and the quality of the reads was checked with FastQC

155   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 (i.e.,

rpoB) and functional genes (i.e., *amoA*_AOA, *amoA*_AOB, and *nxrB*), downloaded from

FunGene [47], were used to quantify the relative abundances and taxonomic affiliations of

160   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}. 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

165   https://github.com/jmtsuji/metannotate-analysis). To allow for approximate between-sample and

between-HMM comparisons, hits for each HMM were normalized both to HMM length and total

number of length-normalized HMM hits for *rpoB* (for more details, see the Github README).

Assembly and binning of metagenome sequence reads

Sequence reads were processed through the ATLAS pipeline (version 2.0.6), which

170   includes quality control, assembly, annotation, binning, bin dereplication, and read mapping [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

8

175

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. Each read required a minimum alignment score ratio of 0.908 (i.e., ~95% sequence identity to the aligned region) to be considered and was mapped to up to 10 best aligned sites across the full genome set. Principal coordinate analysis (PCoA) ordination of samples used Bray-Curtis dissimilarities of relative abundance data using a custom R script.

180 *Analysis of genome bins*

185

190

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 functional genes using reciprocal BLASTp [53] using the BackBLAST pipeline [54], version 2.0.0-alpha2 (doi:10.5281/zenodo.3465955). A concatenated core protein phylogeny, using a set of 74 core bacterial proteins, was generated to assess the phylogenetic placement of *Nitrospira* genome bins. The phylogenetic tree was constructed using IQ-TREE version 1.6.9 [55]. All predicted protein sequences from assembled contigs were compiled and clustered into 99% identity threshold groups within ATLAS; an HMM was used to identify AmoA sequences. *Nitrospira* AmoA were aligned along with reference sequences using MUSCLE [56] and a phylogenetic tree was built using MEGA7 [56]. Cyanase protein sequences (CynS) identified via BackBLAST were further analyzed phylogenetically as described further in the supplemental methods.

9

Data availability

Amplicon sequence data are available on the JGI genome portal under sequencing project

195   ID 1137811. Metagenome sequencing data and high completeness, low contamination genome

bins classified to the *Nitrospirota* phylum were deposited in the European Nucleotide Archive

under study accession number PRJEB30654.

Results and Discussion

Prior to sequence-based analysis of RBC nitrifying biofilm sample communities,

200   ammonia oxidizer cohort sizes were assessed using quantitative PCR (qPCR; Figure 2). The

results showed distinct nitrifying community compositions, both along the treatment train

(spatially; RBC 1 versus RBC 8) and between sampling time points (temporally; 2010 and

2016). Within each RBC train, the measured relative abundance of comammox *Nitrospira* genes,

as well as the proportion of comammox *Nitrospira* within the total community, were consistently

205   higher in RBC 1 than RBC 8 (Figure 2, Table S5), as were ammonium concentrations in samples

taken from directly below the same sampled RBCs (Figure 1C, Table S4). Comammox

Nitrospira amoA genes were more abundant than those of other ammonia oxidizers for most

samples and time points (Figures 2, Figure S1, Table S5). In addition to comammox *Nitrospira*,

the qPCR data also indicated a higher relative abundance of thaumarchaeotal 16S rRNA genes

210   (inferred to represent AOA thaumarchaeota based on other data) in RBC 8 than RBC 1 in the

2010 samples. In contrast, in 2016, thaumarchaeotal 16S rRNA gene abundances were higher in

RBC 1 than RBC 8 for all sampled four trains and were roughly equivalent in relative 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

215   orders of magnitude lower in 2016 than in 2010. Higher resolution sampling of all 2016 RBC

10

stages showed the general trend of a decrease in *amoA* gene abundances across all four RBC flowpaths (Figure S1).

Although 16S rRNA genes cannot be used to differentiate between comammox *Nitrospira* and strict nitrite oxidizers [7], the overall relative abundance patterns of *Nitrospira*

220   amplicon sequence variants (ASVs; Figure 3) were consistent with observations from qPCR data (Figure 2), showing higher relative abundances of *Nitrospira* in RBC 1 compared to RBC 8 for each sample pair. Eight ASVs present at ≥1% relative abundance were classified as *Nitrospira* (Figure 3, Supplemental file 1) and, taken together, these ASVs dominated all other populations detected in the RBC biofilm samples. In addition to *Nitrospira*, five ASVs present at ≥1%

225   relative abundance were classified into the *Nitrosomonadaceae* family (Figure 3), and four corresponding to the genus *Nitrosomonas* affiliated only with the 2010 samples. The fifth *Nitrosomonadaceae* ASV affiliated with 2016 samples (ASV ID 4), but 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 and phylogenetic distance from canonical AOB (data

230   not shown). Thus, qPCR data were consistent with ASV data, with fewer AOB detected in 2016 samples overall. We detected a single thaumarchaeotal ASV that affiliated with the known AOA population within the RBCs (*Nitrosocosmicus* genus; [29]), 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).

235   Single copy taxonomic marker gene analysis of unassembled metagenome sequence data was consistent with qPCR and 16S rRNA gene sequence data, revealing that *Nitrospira* sequences dominated all samples and comprised between 8 and 32% of the total *rpoB* gene sequences among RBC samples from the two sampling years (Figure 4). The *Nitrospira-*

240   affiliated *rpoB* gene relative abundances were higher in RBC 1 than RBC 8 for each train

sampled. In addition, the normalized relative abundances of comammox *Nitrospira amoA* genes

correlated well with the relative abundances calculated from qPCR data for the proportion of

their *amoA* genes within the RBC biofilm communities (Spearman's rank correlation, $r_s = 0.79$, *p*

< 0.001).

Normalized relative abundances of functional genes detected in unassembled read data

245   allowed us to further estimate the relative community contributions of nitrifying community

members. Assuming one *amoA* gene copy per genome, comammox *Nitrospira* represented

between ~5-30% of the total microbial community for all samples based on normalized *amoA*

HMM hits (Figure 5B). *Nitrospira*-associated *rpoB* and normalized *amoA* gene abundances were

similar (Figures 4, 5B), suggesting that the majority of *Nitrospira* present in the RBCs were

250   comammox bacteria. The HMM search for *nxrB* revealed that sequences affiliated with this

model were also prevalent in RBC biofilm metagenomes, with combined hits at over 50%

abundance normalized to total *rpoB* hits (Figure 5C). This normalized relative abundance is

likely an overestimate of the true community contribution of *nxrB*, because the closest BLASTP

search result for some non-*Nitrospira* affiliated *nxrB* HMM hits was to a related protein (e.g.,

255   dimethyl sulfoxide reductase; data not shown). Nevertheless, 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 the dominance of

comammox or NOB *Nitrospira* in the RBCs. Normalized relative abundances of *Nitrospira*-

affiliated *nxrB* sequences that exceed the corresponding relative abundances of *Nitrospira*-

260   affiliated *rpoB* sequences may be due to multiple copies of *nxrB* common to *Nitrospira* genomes

12

[57], in contrast to single copies common for *rpoB* genes [58], or may be due to known biases that occur during HMM hit count normalization.

Consistent with qPCR data, *Nitrosomonas*-affiliated *rpoB* gene sequences were detected at ≥1% in the 2010 samples only. The *rpoB* gene of AOA was not detected at ≥1% relative

265 abundance in any of the samples but ranged from 0.01-0.3% across all samples (data not shown). The HMM search for the *amoA* gene of AOA similarly 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

270 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, but the AOA *amoA* normalized relative abundance pattern was nevertheless 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

275 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 the AOB *Nitrosospira* were also

280 detected, but always at <1% relative abundance. Overall, the relative abundance patterns for AOB *amoA* genes correlated well with their corresponding relative abundances determined by qPCR for the proportion of their *amoA* genes within the total community (Spearman's rank correlation, $r_s = 0.97$, $p < 0.001$).

13

Metagenome-assembled genomes (MAGs) obtained from RBC metagenomic data were

285   also examined for the temporal and spatial distributions of *Nitrospira* populations. Summed

together, *Nitrospira* MAGs (see taxonomic classification description below) had a higher relative

abundance in RBC 1 than RBC 8 for each sample pair (Figure 6). This pattern was also the case

for the bins classified as clade A comammox *Nitrospira*, which comprised a majority of recruited

reads to *Nitrospira* MAGs across most samples. At the same time, individual *Nitrospira* MAGs

290   displayed distinct temporal and spatial distributions. Several of the comammox MAGs had

different abundance patterns in 2010 than in 2016. Specifically, RBC035, RBC100, and RBC069

were present at higher relative abundances in 2010 samples than in 2016 samples, and relative

abundances of RBC001 and RBC083 were generally higher in 2016 than 2010 samples (Figure

6, Supplementary file 2). Multiple comammox *Nitrospira* populations were found in the same

295   RBC, and this high diversity of comammox *Nitrospira* contrasts with the observed low AOA

diversity (i.e., only one *Nitrosocosmicus* species; see below) for all RBCs sampled. Although

0.5-3.1% (median 1.1%) of assembled reads associated with *Nitrospira* MAGs were mapped to

more than one MAG (data not shown), this level of crossover read mapping is insufficient to

explain the high observed diversity of comammox *Nitrospira* for most RBCs, meaning that the

300   high observed diversity is likely not an artefact of read mapping. Using all of the dereplicated

bins, the overall microbial communities of the samples group by year and RBC number (Figure

S2B), similar to an ordination prepared from 16S rRNA gene sequence data (Figure S2A). This

demonstrates that distinct microbial communities were present in each of the two sampling years

and along the RBC flowpaths.

305   The dominance of comammox *Nitrospira* in biofilm samples collected from RBCs

treating municipal wastewater of Guelph, Ontario, Canada is in contrast to previous studies of

full-scale activated sludge WWTPs that report 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,

310 along with other recent studies [12, 15, 17, 18, 22], shows that comammox *Nitrospira* can dominate municipal 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

315 [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 treatment

320 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, 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

325 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 [59]. 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

15

330   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, 60]. A large surface area for attached growth may be an important factor explaining

335   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, Ontario.

        Although relatively low ammonium concentrations likely contribute to the overall

340   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

345   observed for both sample years (Figures 2, 5, 6). However, ammonium concentrations below

RBC 1 were already relatively low (e.g., <18 µM for all 2016 samples; Figure 1C), and it may be

that ammonium concentrations in RBC 8 influent (e.g., <4 µM for all 2016 samples; Figure 1C)

were below the optimum for comammox *Nitrospira*. Consistent with an "oligotrophic lifestyle",

Kits *et al.* [61] reported a high apparent affinity of *Ca.* Nitrospira inopinata for ammonium

350   ($K_{m(app)}$ 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

16

ammonium of the comammox *Nitrospira* detected in the RBCs are unknown but may be similar those reported for *Ca.* N. inopinata.

355   Ammonium concentrations alone may not explain comammox *Nitrospira* abundances, however. 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

360   seen for AOA [e.g. 62–66], 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 [67].

        In addition to targeting nitrification-associated genes to identify comammox *Nitrospira* in

365   extracted DNA or unassembled metagenomic sequences, we explored *Nitrospira* diversity within assembled and binned metagenome sequence data. From a final dereplicated dataset with 101 genome bins (Supplementary file 2), 15 MAGs were classified within the *Nitrospirota* phylum with high completeness (≥75%) and low contamination (<10%; Table S6). All but one of these *Nitrospirota* MAGs were classified under the Genome Taxonomy Database (GTDB) taxonomy

370   within the *Nitrospira* or *Nitrospira* A genera of the *Nitrospiraceae* family; the final MAG (*Nitrospira* bin RBC003) was classified within the uncultured UBA8639 lineage (genus and family name) of the *Nitrospirales* order. Using a concatenated set of 74 core bacterial proteins, these MAGs were further classified according to traditional *Nitrospira* phylogeny (Figures 7, S3). A majority (10 of 15) of the *Nitrospirota* MAGs clustered within clade A comammox

375   *Nitrospira*, four clustered within sublineage I (strict nitrite-oxidizing bacteria), and one (which

17

was classified to the UBA8639 lineage according to GTDB taxonomy) clustered within

sublineage IV (strict nitrite-oxidizing bacteria) of the *Nitrospira*. Thus, the set of 15 *Nitrospirota*

MAGs will be referred to as *Nitrospira* MAGs for conciseness. This high diversity of *Nitrospira*

MAGs is consistent with multiple *Nitrospira* ASVs detected with 16S rRNA gene amplicon

380   sequencing (Figure 3). Using average nucleotide identity (ANI) analysis, the 15 *Nitrospira*

MAGs were compared to one another and to *Nitrospira* genomes from enrichment cultures and

metagenomic surveys (Figure S3). 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* MAGs had <95% ANI to other reference genomes.

385   This suggests that these MAGs represent distinct comammox *Nitrospira* species from most other

previously described comammox species.

        Based on 99% identity clustering of amino acid sequences, all assembled *amoA* genes of

Nitrospira spp. grouped phylogenetically within comammox *Nitrospira* clade A (Figure S4),

aside from RBC group F (a truncated sequence which was dropped from phylogenetic analysis).

390   This is consistent with other studies reporting comammox *Nitrospira* in WWTPs [7, 12, 14–18].

Clade B comammox *Nitrospira* were not detected and the factors affecting the presence and

absence of clades A and B remain unknown. Relatedness of *Nitrospira* bins based on *amoA*

phylogeny (Figure S4) was similar to that obtained with the concatenated protein set, but not

identical (Figures 7, S3). Recently, Xia *et al.* [14] proposed that clade A could be further

395   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 S4).

18

Translated sequences of clade A.2 *amoA* genes from the RBCs were most closely related to

400   *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-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 (i.e., <95%) for many of

405   the bins (Figure S3), which is further evidence that at least some RBC comammox bacteria and

even the NOB *Nitrospira* represent novel species.

Using a reciprocal BLASTP analysis, the *Nitrospira* MAGs were further evaluated for

the presence of *amo, hao,* and *nxr* genes to assess potential contributions to nitrification [5, 6].

Nitrospira bin RBC047 contained all the genes necessary to perform complete ammonia

410   oxidation, although it was missing an ammonium transporter gene (Figure 7). *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 comammox MAGs that contained the *amoA* gene, which is used as a marker gene

for comammox *Nitrospira*. The other five MAGs that clustered phylogenetically within clade A

415   comammox *Nitrospira* contained other genes for ammonia oxidation, supporting their

phylogenetic placement on the tree. The other *Nitrospira* MAGs contained genes only expected

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 (Figure S3). Overall, one *Nitrospira* MAG

420   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

19

lacked the complete gene pathway. 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

425   oxidation.

High functional diversity was observed in the RBC *Nitrospira* MAGs. In the current study, the presence of a nearly complete *ure* operon (Figure 7) 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*

430   genomes and enrichment cultures that use urea [5, 6, 67, 68]. Because the comammox *Nitrospira* bins did not contain any *fdh* genes for formate dehydrogenase (Figure 7), these RBC populations likely cannot use formate as an alternative electron donor, which can be used by NOB *Nitrospira* and most clade B comammox bacteria [68–71]. Most clade A comammox bacteria described so far lack the genes for formate dehydrogenase [68], with the exception of *Nitrospira* bin RCA

435   [72]. Also consistent with previous findings [18, 68], several of our comammox *Nitrospira* MAGs contained partial pathways for group 3b [Ni-Fe] sulfur-reducing hydrogenase genes (*hyb* and *hyd* genes; Figure 7), and several of these MAGs also contained the genes for a putative group 4 hydrogenase (*hyf*), which indicates that hydrogen may serve as an alternative electron acceptor for some of these comammox *Nitrospira*. Multiple clade A comammox MAGs

440   contained genes for accessory proteins for hydrogenases (*hyp*) but none of the *Nitrospira* MAGs contained group 2a [Ni-Fe] hydrogenase genes (*hup*), as seen in the NOB *Nitrospira* *moscoviensis* [73]. All of the comammox *Nitrospira* MAGs lacked the gene for assimilatory nitrite reduction (*nirA*).

Surprisingly, two clade A comammox *Nitrospira* bins (RBC069 and RBC093) that

445   contained genes for ammonia oxidation also encoded the cyanase gene (*cynS*; Figures 7, 8). Prior

to this observation, only canonical NOB *Nitrospira* were known to possess this gene [68, 71, 72,

74, 75]. The *cynS* genes were on long contigs (i.e., 182 kb and 121 kb, RBC069 and RBC093,

respectively), and the primary sequences of most genes adjacent to the *cynS* genes, when queried

against RefSeq using BLASTP, were most similar to genes belonging to known comammox

450   *Nitrospira*. This implies that the presence of *cynS* in RBC comammox MAGs was not caused by

an error in genome binning. Both cyanase genes were distinct from those of other strict NOB

Nitrospira, sharing <75% identity (blastn) to *Nitrospira* homologues in the non-redundant

nucleotide database. Nevertheless, the predicted primary sequences of the *cynS* genes contained

the key active site residues for cyanases (Arg96, Glu99 and Ser122) proposed previously [76],

455   further supporting the functional role of the enzymes (Figure 8A).

Interestingly, the phylogenetic placement of the commamox cyanases was basal to all

other known cyanases among canonical NOB *Nitrospira* (Figure 8A). This basal placement does

not match housekeeping gene phylogenies of *Nitrospira* (e.g., Figure 7), indicating the potential

for lateral gene transfer for cyanase acquisition. Indeed, we detected two toxin-antitoxin genes in

460   the vicinity of the cyanase gene for RBC093, supporting the potential for a recent lateral gene

transfer event [26]. Bins containing *cynS* together represent a substantial proportion of the RBC

microbial community (i.e., ~10-30%; Figure 8B). If these genes are expressed and produce

functional gene products, our results indicate that *Nitrospira*, including novel *cynS*-containing

comammox *Nitrospira*, dominate cyanate degradation to ammonia within RBC biofilm for a

465   direct supply of ammonia (in the case of comammox *Nitrospira*) or for supplying ammonia to

other nitrifiers through reciprocal feeding [74].

21

The *Nitrospira* in the RBCs lack traditional cyanate transporters (*cynABD*), though 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 *cynD* among certain

470 *Nitrospira* (Figure 7). These hits may represent another ABC transport protein, a nitrate/sulfonate/bicarbonate ABC transporter (accession number WP_121988621). If the *Nitrospira* are unable to transport cyanate, our results may suggest that these *Nitrospira* degrade cyanate generated internally, for example from the degradation of carbamoyl phosphate from the urea cycle or from a spontaneous isomeric change of urea to cyanate [77–80]. However, it is also

475 possible that alternate transporters could be used for uptake of cyanate, such as nitrate permease, as seen in other microorganisms [81]. If confirmed, this would represent an undescribed method of nitrogen uptake and utilization for comammox bacteria.

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

480 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 6), implying that complex physicochemical factors govern comammox in the RBCs.

485 Several comammox *Nitrospira* genome MAGs may even represent closely related strains (Figure S3), adding further interest to understanding the factors influencing the distribution and ecology of these bacteria.

Genome bins of other ammonia oxidizers were also recovered. Following dereplication, four bins were classified as *Nitrosomonas* (Supplementary file 2). These populations were

490   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. Classified as *Ca.* N. exaquare under GTDB

495   taxonomy, the AOA bin had an ANI of 96.1% between it and the *Ca.* N. exaquare genome

(GenBank accession number CP017922.1), which indicated that this AOA MAG corresponded

to the same archaeon enriched and characterized from an RBC biofilm sample.

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 that AOA were

500   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]. Lower ammonium concentrations in recent

years (Figure S5) may have shifted the *Ca*. N. exaquare AOA population toward RBC 1, and

cleaning of the RBC trains in 2014 would have impacted the microbial communities. Lower

505   ammonium concentrations beneath RBCs and RBC cleaning may have impacted AOB

populations 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 [61, 82, 83] and may be outcompeted

when ammonium concentrations are lower by ammonia oxidizers with higher ammonia affinities

510   (i.e., comammox *Nitrospira* and AOA).

23

Conclusions

This study used complementary methods to demonstrate that comammox *Nitrospira* dominate microbial communities in biofilms from this engineered WWTP system. Detected comammox *Nitrospira* exhibited a high level of taxonomic and functional diversity, in contrast

515   to the single population of AOA found 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, 84], which would offer relatively low habitat heterogeneity.

520   Although comammox *Nitrospira* were numerically dominant, their contributions to 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 that lacks cultured representatives [14],

525   along with the *cynS*-containing comammox populations, would be ideal targets for cultivation.

The Guelph WWTP RBCs represent a unique and useful system to study the ecology of comammox *Nitrospira* as a result of inbuilt 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

530   WWTPs discharge into environments that have low assimilative capacities or are ecologically sensitive. Understanding the microorganisms that are present within water treatment systems is an important step towards optimizing operational practices for improved effluent quality in municipal, industrial, and aquaculture-associated water treatment systems.

Acknowledgements

Competing interests

The authors declare no competing interests.

References

1.   Daims H, Lücker S, Wagner M. A new perspective on microbes formerly known as

545   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

550   rRNA and *amoA* sequence analysis: Implications for molecular diversity surveys. *Appl 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

555   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

560   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.

565   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.

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

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.

575   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.

580 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.

585 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
590 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 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
595 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.

600   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

605   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.

610   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 into the ecophysiology of the comammox bacterium "*Candidatus* Nitrospira nitrosa".

615   *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–

28

896.

620   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
625   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
630   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
635   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

29

640   *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.

645   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.

650   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.

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

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.

660   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.

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*

685   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.

690   55.   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.

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

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

695   57.   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.

58.   Dahllof I, Baillie H, Kjelleberg S. *rpoB*-based microbial community analysis avoids

limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Appl Environ Microbiol*

700   2000; **66**: 3376–3380.

59.   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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. Limpiyakorn T, Sonthiphand P, Rongsayamanont C, Polprasert C. Abundance of *amoA* genes of ammonia-oxidizing archaea and bacteria in activated sludge of full-scale wastewater treatment plants. *Bioresour Technol* 2011; **102**: 3694–3701.

33

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

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. Koch H, Galushko A, Albertsen M, Schintlmeister A, Gruber-Dorninger C, Lücker S, et al. Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* 2014; **345**: 1052–1054.

74. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate

34

745   as an energy source for nitrifiers. *Nature* 2015; **524**: 105–108.

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

76.   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

750   formation of the enzyme active site. *Structure* 2000; **8**: 505–14.

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

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

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

80.   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.

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

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

83.   Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: The quest for

765   niche specialisation and differentiation. *Trends Microbiol* 2012; **20**: 523–531.

84.   Probst AJ, Auerbach AK, Moissl-Eichinger C. Archaea on human skin. *PLoS One* 2013;

       8: e65388.

Figure legends

770 **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,

775 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

780 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.

785

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

790 relative abundance, in percent. *Nitrospira* ASVs are highlighted in orange, AOB ASVs are highlighted in green, and the AOA ASV is highlighted in blue.

37

Figure 4 Taxonomic profiling of rotating biological contactor (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*, because 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.

Figure 6 Relative abundance of *Nitrospira* bins across samples based on total recruited reads.

815   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.

Figure 7 Phylogeny and gene pathways in *Nitrospira* bins and reference genomes. The bins

820   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*

825   *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*, nitrite/nitrate

830   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.

835   **Figure 8** Phylogeny and relative abundances of cyanases (CynS) detected in RBC metagenome

sequencing data. (A) Maximum likelihood phylogeny of CynS primary sequences detected in

39

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

840   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 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

845   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 proportion (%) of mapped assembled reads to genome data.

850

855
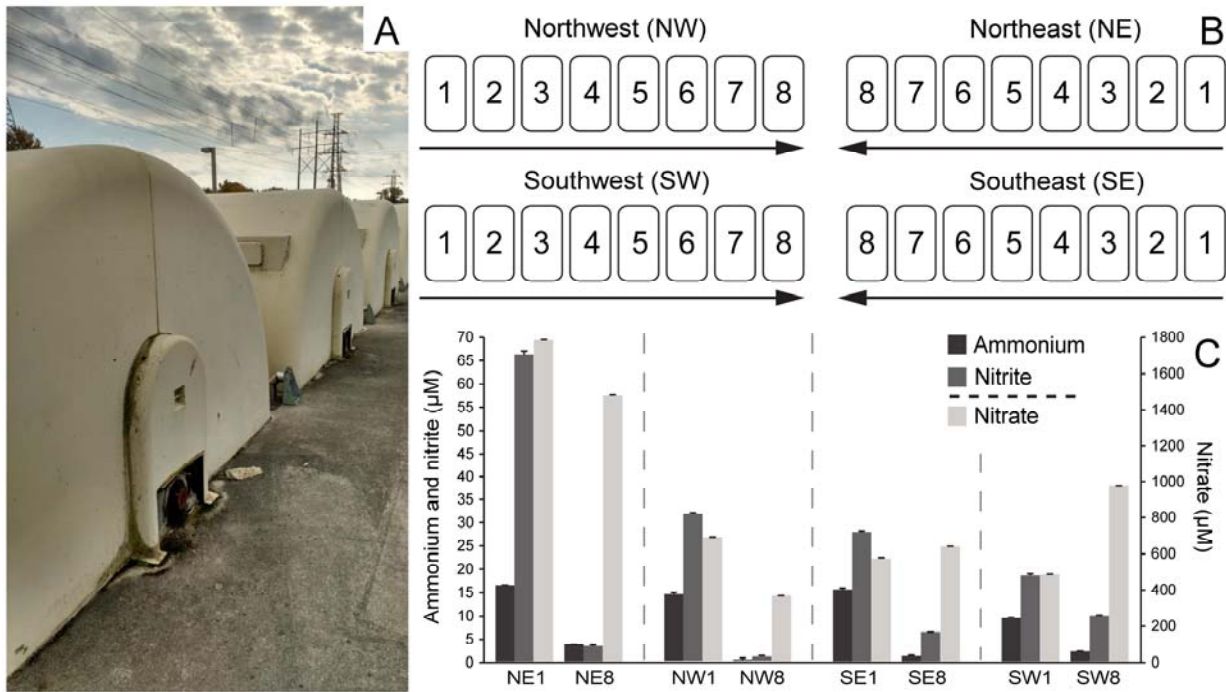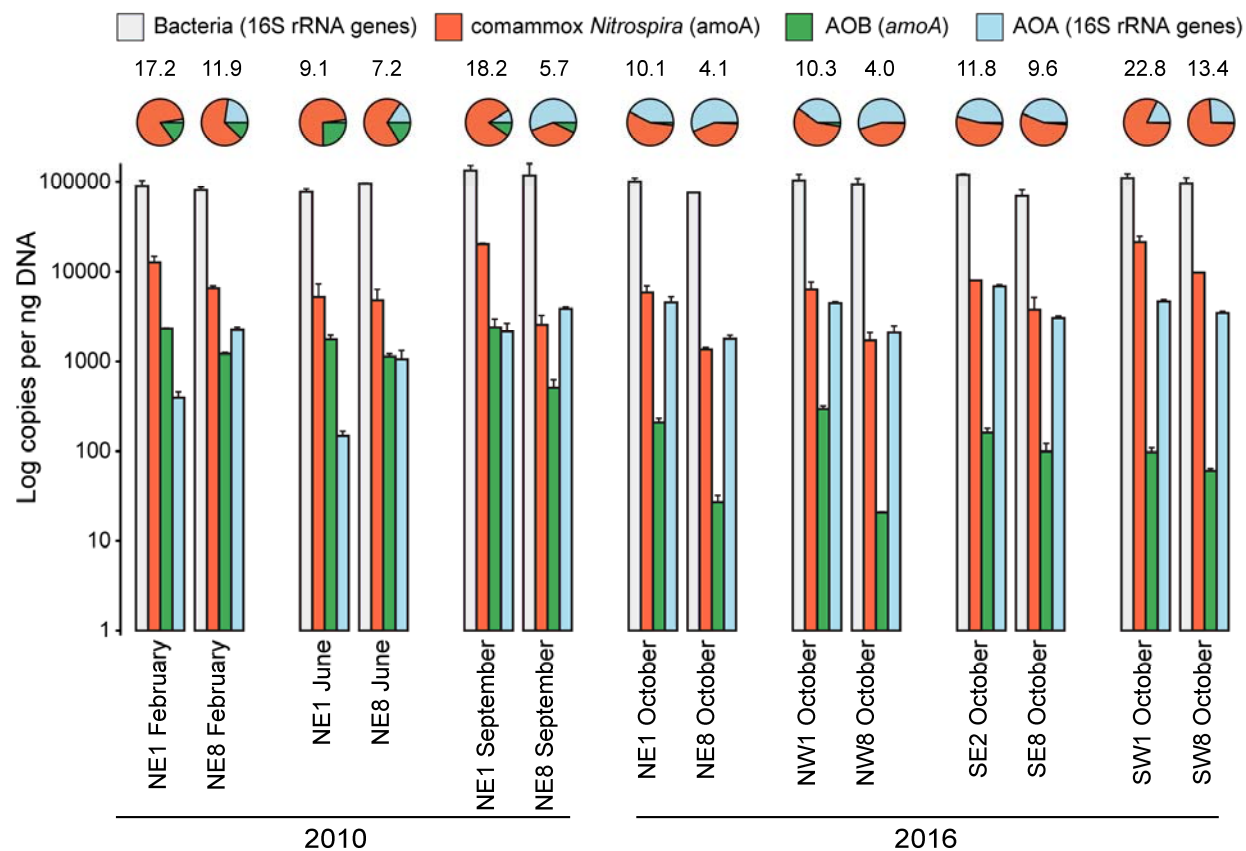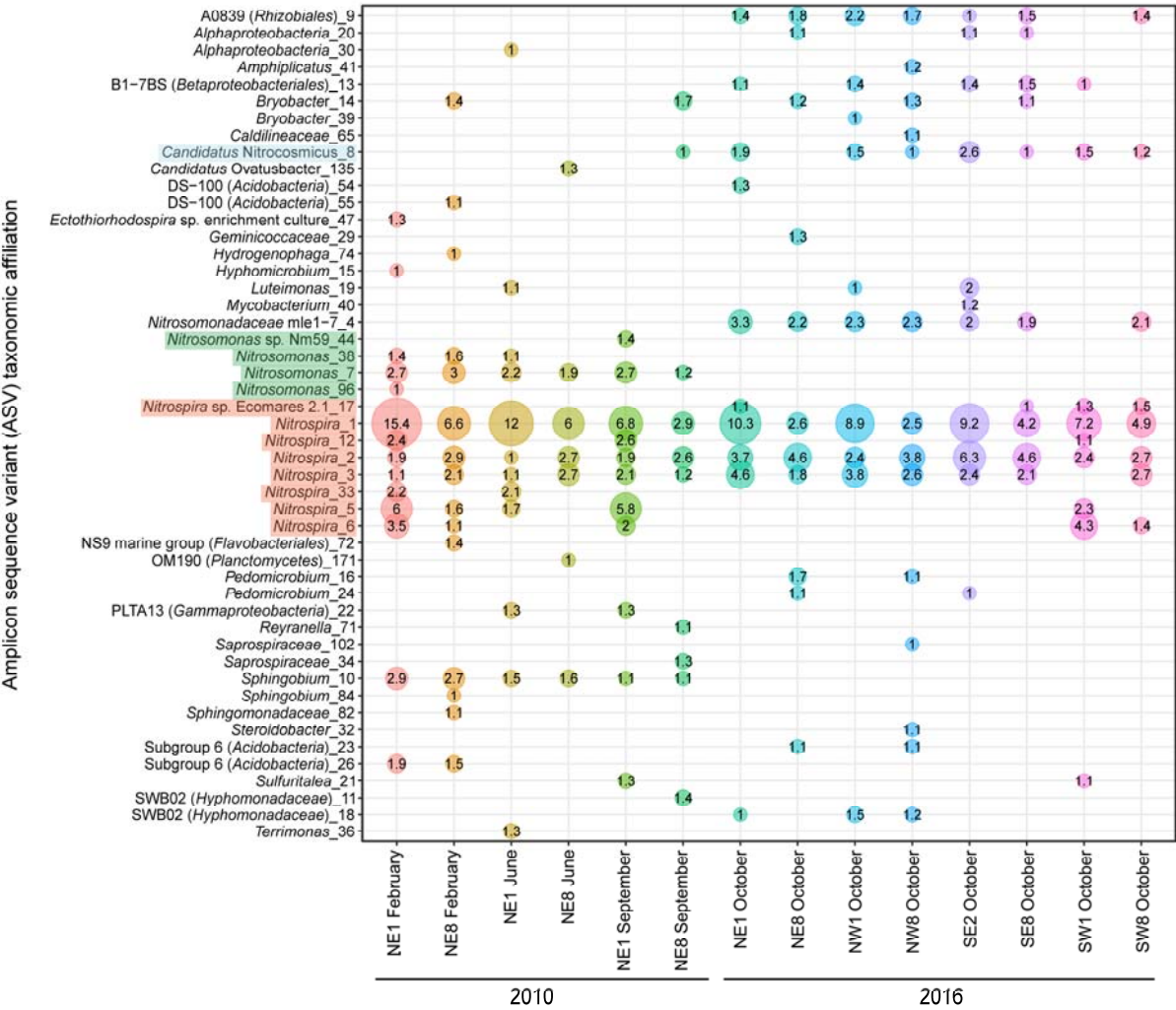
Figures

860

Figure 1

865

Figure 2

880

Figure 3

885

43

890   **Figure 4**

895

44

900   **Figure 5**

Figure 6

905

910

Figure 7

Figure 8

48

915