Genome-enabled insights into the ecophysiology of the comammox bacterium *Candidatus* Nitrospira nitrosa

Pamela Y. Camejo[1], *Jorge Santo Domingo*[2], *Katherine D. McMahon*[1,3], *Daniel R. Noguera*[1,*]

[1]Department of Civil and Environmental Engineering, University of Wisconsin - Madison, Madison, WI, USA

[2]Environmental Protection Agency, Cincinnati, OH, USA

[3]Department of Bacteriology, University of Wisconsin - Madison, Madison, WI, USA

Email addresses:

camejo.pamela@gmail.com

Santodomingo.Jorge@epa.gov

trina.mcmahon@wisc.edu

noguera@engr.wisc.edu

* Corresponding author: Daniel R. Noguera, 1415 Engineering Drive, Madison, WI 53706.

Email: noguera@engr.wisc.edu; Tel: 608-263-7783; Fax: 608-262-5199

ABSTRACT

The recently discovered comammox bacteria have the potential to completely oxidize ammonia to nitrate. These microorganisms are part of the *Nitrospira* genus and are present in a variety of environments, including Biological Nutrient Removal (BNR) systems. However, the physiological traits within and between comammox- and nitrite oxidizing bacteria (NOB)-like *Nitrospira* species have not been analyzed in these ecosystems. In this study, we identified *Nitrospira* strains dominating the nitrifying community of a sequencing batch reactor (SBR) performing BNR under micro-aerobic conditions. We recovered metagenomes-derived draft genomes from two *Nitrospira* strains: (1) *Nitrospira* sp. UW-LDO-01, a comammox-like organism classified as *Candidatus* Nitrospira nitrosa, and (2) *Nitrospira* sp. UW-LDO-02, a nitrite oxidizing strain belonging to the *Nitrospira defluvii* species. A comparative genomic analysis of these strains with other *Nitrospira*-like genomes identified genomic differences in *Ca.* Nitrospira nitrosa mainly attributed to each strains' niche adaptation. Traits associated with energy metabolism also differentiate comammox from NOB-like genomes. We also identified several transcriptionally regulated adaptive traits, including stress tolerance, biofilm formation and micro-aerobic metabolism, which might explain survival of *Nitrospira* under multiple environmental conditions. Overall, our analysis expanded our understanding of the genetic functional features of *Ca*. Nitrospira nitrosa, and identified genomic traits that further illuminate the phylogenetic diversity and metabolic plasticity of the *Nitrospira* genus.

INTRODUCTION

Nitrification is a microbiological process that plays an important role in the nitrogen (N) cycle. This process has been conventionally known as a two-step reaction. The first step, oxidation of ammonia to nitrite, is performed by ammonia-oxidizing bacteria (AOB) or archaea (AOA) and the second step, oxidation of nitrite to nitrate, is carried out by nitrite-oxidizing bacteria (NOB). Recently, the discovery of a new player with the potential to completely oxidize ammonia to nitrate, as in the case of complete ammonia oxidizing (comammox) organisms,[1, 2] has dramatically changed our understanding of microbial mediated N transformations in engineered and natural systems.

Comammox bacteria have been classified within the genus *Nitrospira*. Members of this genus were conventionally regarded as NOB and were thought to rely only on nitrite for growth. However, the genomes of the four comammox-like *Nitrospira* identified to date (*Ca.* Nitrospira nitrosa, *Ca.* Nitrospira nitrificans, *Ca.* Nitrospira inopinata and *Nitrospira* sp. Ga0074138[1-3]), encode the genes necessary for ammonia and nitrite oxidation, suggesting that *Nitrospira* are much more metabolically versatile organisms. Furthermore, comammox-like *Nitrospira* have been identified in a variety of habitats, including groundwater wells, drinking water biofilters, wastewater treatment plants (WWTPs) and other soil and aquatic environments.[4] These findings have prompted questions regarding the ecological significance and lifestyle of these organisms in each of these ecosystems.

Nutrient removal in WWTPs relies on nitrifying organisms to remove N from the wastewater. *Nitrospira*-like bacteria appear to be the dominant nitrite-oxidizers[5-7] in most WWTPs and laboratory scale reactors. The abundance of comammox in WWTPs has been briefly surveyed,

and preliminary results show this functional group is present in these systems.[4] However, genetic and functional adaptations of comammox to this environment have not been addressed.

In this study, the community performing N removal in a Biological Nutrient Removal (BNR) lab-scale reactor was analyzed to explore the genomic basis for comammox ecophysiology. A sequencing batch reactor (SBR) was operated under cyclic anaerobic and micro-aerobic conditions using two different operational conditions. During the first stage (nitrite addition during a micro-aerobic phase), two *Nitrospira*-like strains were enriched in the reactor. Draft genome sequences of these two strains were assembled from metagenomic data; one of them was identified as a commamox organism, the other as an NOB. Here, we used the draft genomes of these strains, as well as genomes from both NOB- and comammox-related bacteria, to perform a comparative genome analysis of the genus *Nitrospira*.

MATERIAL AND METHODS

Operation of Lab-Scale Sequencing Batch Reactor

A laboratory-scale SBR was originally inoculated with activated sludge obtained from the Nine Springs WWTP in Madison, WI, which uses a modified University of Cape Town (UCT) process designed to achieve biological P removal[8] and operates with high aeration rates.[9] Synthetic wastewater containing acetate as the sole carbon source was used for the feed, as described elsewhere.[10] The hydraulic retention time (HRT) and solids retention time (SRT) were 24 h and 80 days, respectively. The pH in the system was controlled to be between 7.0 and 7.5.

The 2-liter reactor was operated under alternating anaerobic and low oxygen cycles. During Stage 1 of operation the cycles consisted of 2 h anaerobic, 5 h micro-aerobic, 50 min settling and 10 min decanting. At the beginning of the micro-aerobic phase, sodium nitrite was added to reach

an in-reactor concentration of 10 mg N-NO_2^-/L and stimulate the use of nitrite as an electron donor. In addition, an on/off control system was used to limit the amount of oxygen pumped to the reactor (0.02 L/min) and maintain low dissolved oxygen (DO) concentrations in the mixed liquor, as described elsewhere.[10] After 100 days of operation, the nitrite supplement was eliminated and the reactor cycle was changed to: 1.5 h anaerobic, 5.5 h micro-aerobic, 50 min settling and 10 min decanting (Stage 2).

Sample Collection and Analytical Tests

To monitor reactor performance, mixed liquor and effluent samples were collected, filtered through a membrane filter (0.45 μm; Whatman, Maidstone, UK) and analyzed for acetate, PO_4^{3-}–P, NH_4^+-N, NO_3^--N, and NO_2^--N. The concentrations of PO_4^{3-}–P were determined according to Standard Methods.[11] Total ammonia (NH_3 + NH_4^+) concentrations were analyzed using the salicylate method (Method 10031, Hach Company, Loveland, CO). Acetate, nitrite and nitrate were measured using high-pressure liquid chromatography as previously described.[10]

Biomass samples from the reactors were collected weekly and stored at -80°C until DNA extraction was performed. DNA was extracted using UltraClean® Soil DNA Isolation Kit (MoBIO Laboratories, Carlsbad, CA). Extracted DNA was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and stored at -80°C.

Metagenome Sequencing, Assembly and Binning

Samples from day 100 (Stage 1) and days 317, 522 and 674 (Stage 2) were selected for metagenomic analysis. Illumina TruSeq DNA PCR free libraries were prepared for DNA extracts according to the manufacturer's protocol and paired-end sequenced on either the Illumina HiSeq 2000 platform (v4 chemistry, 2 × 150 bp; 522-day sample), or the Illumina MiSeq platform (v3

chemistry, 2 × 250 bp; other samples). Unmerged reads were quality-trimmed and filtered with Sickle (https://github.com/ucdavis-bioinformatics/sickle.git), using a minimum phred score of 20 and a minimum length of 50 bp. The metagenomic reads from the 100-day sample (Stage 1) were assembled using IDBA-UD.[12] Individual genome bins were extracted from the metagenome assembly with the R package 'mmgenome'[13] using the differential coverage principle.[14] The bins were initially extracted by plotting the genome coverage of contigs in metagenomes from day 100 and 317. During the bin extraction, GC content and taxonomy of contigs were also taken into consideration.

After binning, SSPACE was used to filter small scaffolds (length < 1,000 bp), extend scaffolds, and to fill gaps.[15] Genome completeness and contamination was estimated using CHECKM 0.7.1.[16] Table S1 displays quality metrics of the draft genomes after each of the steps previously described. Two putative *Nitrospira*-like bins were identified and annotated using MetaPathways v 2.0.[17] To further reduce contamination in these assembled bins, scaffolds with Open Reading Frames (ORFs) having less than 85% nucleotide identity and 0% protein identity to other *Nitrospira* genomes were removed from the bins.

Average Nucleotide Identity (ANI)

Pair-wise ANI values of *Nitrospira*-like genomes were obtained using the ANIm method[18] and implemented in the Python script 'calculate_ani.py' available at https://github.com/ctSkennerton/scriptShed/blob/master/calculate_ani.py.

Phylogenetic Analyses

The phylogeny of the draft genomes was assessed by constructing a phylogenetic tree using a concatenated alignment of marker genes. First, PhyloSift v 1.0.1[19] was used to extract a set of 38

marker genes from each genome. Then, the extracted marker protein sequences were concatenated into a continuous alignment to construct a maximum-likelihood (ML) tree, using RAxML v 7.2.8.[20] RAxML generated 100 rapid bootstrap replicates followed by a search for the best-scoring ML tree.

For phylogenetic analyses of ammonia monooxygenase subunit A (*amoA*), hydroxylamine reductase (*hao*) and nitrite oxido-reductase subunit A (*nxrA*) genes, full nucleotide datasets were downloaded from the NCBI GenBank database.[21] Alignment was performed on sequences retrieved from the NCBI and the draft genomes using the 'AlignSeqs' command in the DECIPHER "R" package.[22] Phylogenetic trees were calculated using neighbor-joining criterion with 1,000 bootstrap tests for every node, using the MEGA6 software package.[23] Trees were visualized with the assistance of TreeGraph.[24]

Population Structure by Metagenomic Analysis

To estimate the abundance of currently known ammonia oxidizers, comammox, and nitrite oxidizers in the reactor over time, paired-end DNA reads from the metagenomic datasets (days 100, 317, 522 and 674) were competitively mapped to the published genome sequence of 14 AOB (*Nitrosomonas*, *Nitrosospira* and *Nitrosococcus* genera), 6 AOA (*Nitrososphaera*, *Nitrosoarchaeum* and *Nitrosopumilus*), 5 NOB (*Nitrospira* and *Nitrobacter* lineage), 5 anaerobic ammonia-oxidizing (anammox) bacteria (*Ca.* Brocadia fulgida, *Ca.* Brocadia caroliniensis, *Ca.* Kuenenia stuttgartiensis, *Ca.* Brocadia sinica and *Ca.* Jettenia caeni), 4 comammox bacteria (*Ca.* Nitrospira nitrosa, *Ca.* Nitrospira nitrificans, *Ca.* Nitrospira inopinata and *Nitrospira* sp. Ga0074138) and the two *Nitrospira*-like draft genomes retrieved from the reactor, using the software package BBMap version 35.85 (https://sourceforge.net/projects/bbmap). A list of the genomes included in this analysis and the number of reads mapping to each sequence is found in

Table S2. For each organism, the number of unambiguous reads (best hit) mapping to the genomic sequence with a minimum alignment identity of 90% was quantified and normalized by the number of reads in each metagenome, paired-end reads average length and genome size (Table S2).

Orthologous Genes Clusters

To assess the degree of homology in the proteomes of the two *Nitrospira*-like genomes, orthologous genes clusters (OCs) were determined using OrthoMCL.[25] OrthoMCL was run with a BLAST E-value cut-off of 1e-5, and an inflation parameter of 1.5.

Protein products of each ortholog set were compared against the Eggnog database[26] using the eggNOG-mapper tool (https://github.com/jhcepas/eggnog-mapper) (*e*-value $<10e^{-2}$) to determine the COG functional and super-functional category to which they belong.

Accession Numbers

Raw reads and draft genome sequences have been submitted to NCBI and are accessible under the BioProject identifier PRJNA322674.

RESULTS AND DISCUSSION

Nutrient Removal in Lab-Scale Reactor

Results from a typical cycle of the lab-scale SBR at steady-state operation during Stage 1 are shown in Fig. 1. During the anaerobic phase, acetate added at the beginning of the anaerobic phase was completely consumed within an hour (Fig. 1B). P release to the mixed liquor during this condition was not observed (Fig. 1B), indicating the absence of polyphosphate accumulating organisms (PAO) in the reactor. Denitrification was incomplete, with only ~ 60% of the nitrate removed in the anaerobic phase (Fig. 1A), even though the reactor received acetate in this phase. This suggests that efficient acetate uptake was likely performed by glycogen accumulating

177 organisms (GAO), without affecting P concentrations.[27] In addition, nitrite production during this
178 phase (~ 10%) is an indicator of partial denitrification (Fig. 1A).

179 Complete nitrification occurred in the subsequent micro-aerobic stage, where 91% ± 4% of
180 ammonia and 91% ± 8% of the added nitrite were removed (Fig. 1A), with nitrate accumulation
181 accounting for 50% ± 2% of the oxidized nitrogen. During the period of active nitrification, the DO
182 remained below 0.05 mg/L (Fig. 1C), as the oxygen supplied balanced the oxygen uptake rate. The
183 oxygen uptake rate decreased after nitrification ceased, and correspondingly, DO increased. To
184 maintain a low-DO environment, aeration was stopped when DO exceeded the set point (0.2 mg
185 O_2/L) and resumed when DO decreased below the set point. This operation effectively maintained
186 DO below 0.4 mg/L (Fig. 1C).

187 After 100 days of reactor operation under these conditions, the operational parameters were
188 changed to promote simultaneous nitrification, denitrification and P removal (Stage 2). Unlike
189 Stage 1, no nitrite was added to the reactor during micro-aerobic conditions. During this second
190 stage, acetate added at the beginning of the anaerobic phase was used by PAOs for P-cycling, and
191 nitrite and nitrate produced by ammonia oxidization were used as electron acceptors by PAOs
192 during micro-aerobiosis, achieving simultaneous removal of N and P. Results of this stage were
193 described elsewhere.[10]

194 ***Nitrospira*-like Genome Binning**

195 Using a combination of bidimensional coverage and tetranucleotide frequency, two
196 *Nitrospira*-like draft genomes were assembled from a sample collected at the end of Stage 1. The
197 two draft genomes (*Nitrospira* sp. UW-LDO-1 and UW-LDO-2) had 3.9 and 3.5 Mbp in total with
198 average GC content of 54.9% and 59.2%, respectively (Table S1). The reconstructed genomes

were assessed to be nearly complete (completeness ≧90%) with low contamination (≦5%), according to the presence of 43 single-copy reference genes (Table S1).

Since the composite genomes did not encode complete 16S rRNA genes, the average nucleotide sequence identity (ANI) between the draft genomes assembled herein and formerly published *Nitrospira*-like genomes, was used to determine whether UW-LDO-1 and UW-LDO-2 represented distinct species, as this method has been shown to correlate well with previously defined 16S rRNA gene species boundaries.[28] The calculated ANI and fraction of alignment for the *Nitrospira* genomes (Fig. 2) showed that UW-LDO-01 is a representative of the *Ca. Nitrospira nitrosa* species (ANI >94%, Fraction Aligned = 74.9%), while UW-LDO-02 had the closest nucleotide identity to *Nitrospira defluvii* (ANI = 92.4%, Fraction Aligned = 72.4%). None of the other ANI values were greater than 88%, indicating that the two genomes were different from each other, and supports their classification as *Nitrospira nitrosa* UW-LDO-01 and *Nitrospira defluvii* UW-LDO-02, respectively.

Phylogenetic Analysis

A genome tree constructed from a concatenated protein alignment of 38 universally distributed single-copy marker genes[29] confirms the affiliation of *Nitrospira* sp. UW-LDO-01 and UW-LDO-02 with *Ca.* Nitrospira nitrosa and *Nitrospira defluvii*, respectively (Fig. 3). Consistent with this phylogeny, UW-LDO-01 encoded the *amoCAB* operon, responsible for ammonia oxidation, while UW-LDO-02 did not.

The *amoA* and *hao* genes are functional genes involved in redox nitrogen transformations and are also considered phylogenetic markers to study the diversity of ammonia oxidizing microorganisms (AOM).[30-33] The phylogenetic tree topologies based on these genes (Fig. S1)

221 further confirm the classification of UW-LDO-01 as related to *Ca.* Nitrospira nitrosa, although
222 two paralogs of the *amoA* gene are present in the genome of *Ca. Nitrospira nitrosa*, while only
223 one *amoA* gene was found in UW-LDO-1.

224 In addition, the *Nitrospira* sp. UW-LDO-1 and UW-LDO-2 genomes encoded the key
225 enzyme for nitrite oxidation, *nxr*, which can also be used as a phylogenetic biomarker. UW-LDO-
226 1 encoded two paralogs of the periplasmic NXR enzyme while *Nitrospira* UW-LDO-2 only
227 encoded one copy. The affiliation of UW-LDO-1 with *Ca.* Nitrospira nitrosa and UW-LDO-2 with
228 *N. defluvii* was supported by phylogeny based on the *nxrA* gene sequence (Fig. S2), consistent
229 with the phylogenetic analysis of *amoA* and *hao*.

230 **Nitrifying Prokaryotes in Lab-Scale Reactor**

231 The metagenomic analysis of the Stage 1 sample, which corresponds to the operational
232 conditions when nitrite and ammonia were both present under micro-aerobic conditions, did not
233 result in the assembly of any other genome of nitrifying microorganisms. Thus, in order to assess
234 the relative abundance of other known nitrifying prokaryotes present in the reactor, we mapped
235 metagenomic reads to published genomes of comammox, anammox, AOB, AOA and NOB,
236 including *Nitrospira* sp. UW-LDO-01 and UW-LDO-02 (Fig. 3). After a competitive mapping of
237 short-reads from metagenomic samples to each genome (>90% identity), the number of mapping
238 reads was normalized to both metagenome size and reference genome size, and used as a proxy of
239 genome abundance.

240 The metagenomic data show little evidence of AOA and anammox bacteria during the two
241 stages (0.06% and 0.25% of mapping reads, respectively). AOB were detected in the system, albeit
242 representing a small fraction of the community (0.17% and 0.05% of total number of reads during
243 Stage 1 and 2, respectively). Noticeably, *Nitrospira*-like sequences (including comammox- and

NOB-like genomes) recruited the greatest number of metagenomic reads (14.0% of total number of reads) in the Stage 1 sample (Fig. 4A). Within this genus, *Nitrospira sp.* UW-LDO-01 retrieved 32.3% of the reads competitively mapping to the *Nitrospira*-like genomes (Fig. 4B). The published genome of *Ca.* Nitrospira nitrosa retrieved 2.3% of the reads, while less than 4% mapped to other comammox genomes. Therefore, with only a small fraction of reads mapping to other ammonia oxidizers, we propose that *Nitrospira* sp. UW-LDO-01 was the main comammox in the reactor, and the main contributor to ammonia oxidation during Stage 1.

Nitrospira sp. UW-LDO-02 appeared to be the most abundant NOB in the reactor during Stage 1, retrieving 42.4% of the *Nitrospira*-like reads (Fig. 4B), although a large fraction of reads competitively mapping to *N. defluvii* may indicate the presence of other nitrite oxidizing strains in the reactor. Therefore, the nitrite oxidation activity in the reactor was carried out by comammox and NOB.

The metagenomic analysis of Stage 2 samples reveals an overall decrease in the relative abundance of nitrifying organisms in the reactor after transitioning to this operational configuration (Fig. 4A). During this stage, metagenomic reads mapping to NOB and comammox genomes (including *Nitrospira* sp. UW-LDO-01 and UW-LDO-02) decreased to less than 1% of the total number of reads. This was in part due to the removal of nitrite addition during Stage 2. However, the decrease in comammox did not correspond to an increase in the number of reads mapping to other known AOM (Fig. 4A), suggesting the presence of still unrecognized AOM in reactors operated with low DO conditions, as previously reported.[34]

Differential Gene Content Among *Ca.* Nitrospira nitrosa Genomes

Since *Nitrospira* sp. UW-LDO-01 is the second comammox genome representative of *Ca.* Nitrospira nitrosa, and the first comammox genome recovered from a nutrient removal bioreactor,

a comparative analysis of its genetic content was carried out. First, a comparison of gene content among *Ca.* Nitrospira nitrosa (CZQA00000000) and *Nitrospira* sp. UW-LDO-01 was conducted by BlastP comparison of the translated coding DNA sequence (CDS) set, clustering of ortholog proteins, and annotation of representatives of each ortholog cluster (OC) and genome-unique CDSs.

Overall, sequencing and annotation of the UW-LDO-01 genome revealed a genomic inventory highly similar to the genome of *Ca.* Nitrospira nitrosa.[2] The two genomes shared 81.2% of the OCs (3,202 OCs), with UW-LDO-01 and *Ca.* Nitrospira nitrosa having 740 and 857 unique OCs, respectively (Fig. 5A).

OCs belonging to the shared and genome-specific fractions of the two genomes were classified according to their predicted functional role (Fig. 5B-C). 2,906 and 3,009 OCs, in the UW-LDO-01 and *Ca*. Nitrospira nitrosa genomes, respectively, were categorized into functional Clusters of Orthologous Groups (COG) categories. The majority of OCs belonged to the group of "Unknown Function (S)" (23% of shared OCs, and 19% and 23% of the genome-specific fraction in UW-LDO-01 and *Ca.* Nitrospira nitrosa, respectively), indicating a large set of metabolic features not yet elucidated (Fig. 5B-C).

Besides the genes of unknown function, OCs within this shared fraction were mostly represented by the "Cell wall/membrane/envelope biogenesis (M)" (7%), "Signal transduction mechanisms (T)" (6%) and "Energy production and conversion (C)" (6%) families, indicating general conservation of energy metabolism and regulatory mechanisms.

In both genomes, the functional group "Replication, recombination and repair (L)" was over-represented within the genome-specific fraction. This functional category includes transposases, integrases and other mobile genetic elements, and their extensive representation in the genome-

specific fraction indicates that horizontal gene transfer has likely played a significant role in the diversification of *N. nitrosa* strains. The greatest difference among the genome-specific fractions of UW-LDO-01 and *Ca*. Nitrospira nitrosa was the proportion of OCs represented by the "Cell wall/membrane/envelope biogenesis (M)" and "Intracellular trafficking, secretion, and vesicular transport (U)" functional families (Fig. 5C). Glycosyltransferases[35-38], Type IV[39-41] secretion system proteins and other enzymes involved in polysaccharide formation (main component of the biofilm matrix) were enriched within these COG categories in the genome-specific fraction of *Ca.* Nitrospira nitrosa. Differences in biofilm formation capabilities between these strains may relate to specific niche adaptation: *Ca.* Nitrospira nitrosa was enriched in a biofilm, whereas UW-LDO-01 was found in a planktonic habitat in wastewater. Analogous findings have been observed in other genera, where differences among biofilm formation capabilities within the same genus were linked to the genome content of different strains.[42-44] Similar to the results presented here, these genetic differences included the presence of type IV secretion systems and enzymes involved in protein glycosylation.

The comparative genomic analysis also indicated a higher proportion of gene clusters associated with "Lipid transport and metabolism (I)" in *Nitrospira* sp. UW-LDO-01 (Fig. 5C). Genes related to ß-oxidation of long-chain fatty acids to acetyl-CoA were present in the genome of UW-LDO-01, but absent in *Ca*. Nitrospira nitrosa. These genes include a long-chain fatty acid-CoA ligase, acyl-CoA dehydrogenase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase and acetyl-CoA acetyltransferase. Presence of these lipid-related metabolic genes in other *Nitrospira* strains was confirmed, although the complete pathway is lacking in *Nitrospira defluvii*, *Ca.* Nitrospira nitrificans and *Nitrospira* sp. Ga0074138. This feature may represent a competitive advantage of some *Nitrospira* strains in habitats rich in long-chain fatty acids, such as WWTPs.[45]

313 **Metabolic Features in *Nitrospira* Genomes**

314 To explore the diverse metabolic capabilities and provide insights into the common and
315 unique metabolic features encoded in the genome of NOB- and comammox-like strains, we
316 compared the gene inventory of 9 complete and draft genomes classified as *Nitrospira*. The
317 analysis was focused on traits associated to energy production, which are summarized in Table S3.

318 In agreement with previous analyses, only comammox-like genomes encoded ammonia
319 monooxygenase (*amoCAB*) and hydroxylamine dehydrogenase (*haoAB-cycAB*) gene clusters,
320 responsible for ammonia oxidation to nitrite (Table S3), reflecting the capability of this novel
321 *Nitrospira* sub-lineage to perform full-nitrification from ammonia to nitrate.

322 Analysis of nitrite reducing genes revealed that all *Nitrospira* strains encoded a copper-
323 containing dissimilatory nitrite reductases (*nirK*), which catalyzes the nitrite reduction to nitric
324 oxide, a key step in the denitrification process. Despite the widespread presence of this enzyme
325 across the *Nitrospira* genus, former studies have documented no activity of this protein in NOB-
326 like[46] or comammox-like strains[1], where N loss caused by formation of gaseous compounds was
327 not observed. Since it has been predicted that the NXR complex of *Nitrospira* can reduce nitrate
328 to nitrite,[46] these microorganisms appear genetically capable of converting nitrate (product of
329 nitrification) to nitric oxide. Additional experiments are still needed to obtain more insights into
330 this *Nitrospira* trait. Other denitrification genes, such as nitrate reductase (*nar*), nitric oxide
331 reductase (*nor*) or nitrous oxide reductase (*nos*), were not found in the *Nitrospira* strains analyzed
332 here.

333 All comammox-like genomes, including UW-LDO-01, encoded the machinery to hydrolyze
334 urea: the *ureABCDFG* urease operon and the *urtABCDE* urea transport system, suggesting that
335 this *Nitrospira* sub-division possesses a high-affinity uptake system for urea and, thus, is adapted

to habitats where urea is present at low levels. A gene cluster involved in urea metabolism was also found in *N. moscoviensis* (Table S3), although the urea-binding proteins *urtBCDE* were lacking in the genome. Ureolytic activity of both *N. moscoviensis*, *Ca*. Nitrospira nitrosa and *Ca.* Nitrospira nitrificans was formerly tested by incubation of these strains with urea-containing media, where urea hydrolysis to ammonium was observed in both cases.[2, 46] Former studies have also shown the presence of genes for urea utilization in *Nitrospira lenta,*[47] a novel *Nitrospira* species enriched under low temperatures, suggesting that the ureolytic activity might be associated with this lineage. Since several ammonia oxidizers also possess the capability of hydrolyzing urea,[48-50] the presence of this trait in *Nitrospira* might evidence horizontal gene transfer among these functional groups.

A contrasting difference among NOB-like and comammox-like genomes was the capability to convert cyanate into ammonia. Only NOB-like genomes encoded a cyanase hydratase enzyme; former studies have experimentally confirmed cyanate degradation in *N. moscoviensis.*[51] Cyanate is produced intracellularly from urea and carbamoyl phosphate decomposition,[52, 53] and in the environment from the chemical/physicochemical decomposition of urea or cyanide.[54, 55] The presence of a cyanase enzyme benefits nitrite oxidizers because it allows them to detoxify cyanate, and the formed ammonium is then available for assimilation and might also serve as a source of energy for ammonia oxidizers in a process described as "reciprocal feeding".[46, 51] Further experiments analyzing the effect of cyanate in the growth on comammox-like bacteria are needed to understand how cyanate degradation would confer them a biological advantage, besides generation of ammonia.

The analysis also revealed the presence of the gene inventory for the uptake and oxidation of formate, an exclusive feature of NOB. Growth on formate as electron donor has been confirmed

359 in *N. moscoviensis* (under both micro-oxic and anoxic incubations),[46] *Nitrospira japonica*[56] and
360 uncultured *Nitrospira* in activated sludge.[57] Although formate oxidation should be advantageous
361 for organisms thriving in hypoxic or anoxic habitats, which also includes comammox-like bacteria,
362 this feature is absent in the genome of these microorganisms.

363 The genome of *N. moscoviensis* encodes a group 2a [Ni-Fe] hydrogenase (*hupS* and *hupL*)
364 and accessory proteins involved in the maturation and transcriptional regulation of hydrogenases
365 (*hypFCDEAB* and *hoxA*). Furthermore, experiments showed that *N. moscoviensis* was capable of
366 growing by aerobic respiration of H_2.[58] Although the comammox-like genomes lack the subunits
367 of the [Ni-Fe] hydrogenase (Hup), the five genomes analyzed here encoded a group 3 [Ni-Fe]
368 sulfur-reducing hydrogenase gene set (*hydBGDA and hybD*) positioned at the same locus where
369 Hup is located in *N. moscoviensis*. This hydrogenase complex is an heterotetramer with both
370 hydrogenase activity and sulfur reductase activity, which might play a role in hydrogen cycling
371 during fermentative growth.[59] Its beta and gamma subunits, that form the sulfur reducing
372 component, catalyze the cytoplasmic production of hydrogen sulfide in the presence of elemental
373 sulfur. The presence of this complex in the genomes indicates the potential of these
374 microorganisms for oxidizing H_2 using sulfur as electron acceptor, a trait that has not been
375 analyzed in comammox before, but that could confer this sub-group an advantage when growing
376 under anaerobic conditions.

377 Furthermore, the presence of a hyf-like operon (*hyfBCEFGI*), which encodes a putative
378 group 4 hydrogenase complex, was detected in every NOB-like genome, as well as *Ca.* Nitrospira
379 nitrosa, *Nitrospira* sp. Ga0074138 and UW-LDO-01. In *Escherichia coli*, this hydrogenase
380 complex forms part of a second formate hydrogenlyase pathway (oxidation of formate to CO_2 and
381 reduction of $2H^+$ to H_2 under fermentative conditions).[60] This is likely the case for the

hydrogenase-4 present in the genome of NOB-like strains, which co-occur with genes encoding formate dehydrogenase. In comammox, however, the role of this distinct hydrogenase is not as clear. In *Ca.* Nitrospira nitrosa, this complex is found immediately adjacent to a carbon monoxide dehydrogenase (CODH), an enzyme that catalyzes the interconversion of CO and CO_2[61], a genomic feature that would allow this strain to obtain energy from carbon monoxide.[62] Conversely, the genomes of *Nitrospira* sp. Ga0074138 and UW-LDO-01 lack the CODH at this position, which in the case of UW-LDO-01, was confirmed by alignment of the metagenomic reads to this gene. No other neighboring gene of the hydrogenase-4 complex could be associated with this enzyme in these two strains, therefore, the biological role of these genes is still unclear.

Altogether, these results reveal specific traits characterizing the NOB and comammox functional groups: while comammox-like *Nitrospira* has the genomic potential of ammonia and nitrite oxidation and potentially, sulfur reduction, NOB-like strains are distinguished by their cyanate degradation and formate oxidation capabilities; and both urea hydrolysis and H_2 respiration are common traits shared by multiple *Nitrospira* strains.

The Role of Transcriptional Regulation in *Nitrospira*

Transcriptional regulation of gene expression is the most commonly used strategy to control many of the biological processes in an organism, including progression through the cell cycle, metabolic and physiological balance, and responses to environmental stress. This regulation is generally orchestrated by several transcriptional factors (TFs) that directly coordinate the activity of genes by binding to their promoters. Each *Nitrospira*-like genome codes for at least 100 transcriptional regulators, which account for ~3% of the estimated total number of genes, in agreement with TFs in other microorganisms.[63-65] A comparative genomic analysis of full and draft

Nitrospira genomes was used to investigate the repertoire of TFs potentially involved in the survival of these microorganisms under diverse environmental conditions (Table 1).

Among the TFs analyzed, the formate hydrogenlyase transcriptional activator (fhlA)[60, 66] was the only one shared across all the *Nitrospira* genomes, although only NOB-like genomes encode genes of its known regulon, the formate hydrogenase complex. The presence of this transcriptional activator in comammox microorganisms, which appear to be genetically incapable of formate oxidation (Table 1), might represent an ancestral trait shared by *Nitrospira* and lost during diversification. This theory would also support the presence of the group 4 hydrogenase (associated with the formate-hydrogenlyase complex in *E. coli*) in both NOB- and comammox-like groups.

A common feature among some NOB and comammox bacteria is the presence of the transcriptional regulators NhaR[67, 68] and OxyR[69] (Table 1). The first one is associated with stress response to alkaline, acidic, saline and osmotic conditions. OxyR regulates hydrogen peroxide-inducible genes, such as alkyl hydroperoxide reductase (*ahpCF*) and glutaredoxin (*grxA*), encoded in all the *Nitrospira* genomes. Presence of these genes would confer *Nitrospira* an improved fitness advantage over other nitrifying bacteria. For instance, NhaR is lacking in *Nitrosomonas* and *Nitrobacter* and OxyR is not present in *Nitrosomonas* and *Nitrosospira* (based on genome search). Furthermore, the role of NhaR during regulation of *pga* expression,[68] allows the biofilm formation process to be considered as a flexible and dynamic developmental process driven by external conditions, representing another means by which NhaR could promote survival of *Nitrospira*. Likewise, the presence of the chemotaxis regulator CheZ in *Nitrospira* suggests chemotaxis as another important mechanism by which these microorganisms efficiently and rapidly respond to changes in the chemical composition of their environment.

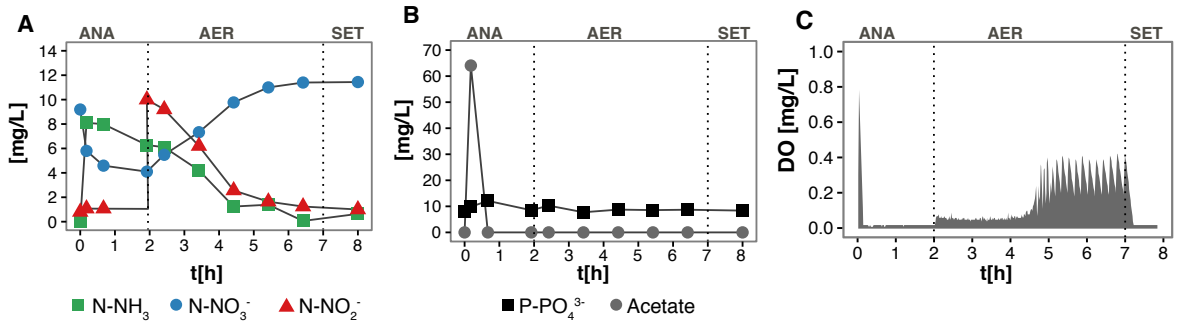
To date, the role of the Fnr-type regulatory protein in *Nitrospira* has not been determined. In other microorganisms, Fnr is part of the signaling involved in the adaptation to micro-oxic environments,[70-74] where it acts as oxygen sensor and regulator of genes involved in anaerobic and micro-aerobic metabolism. In *Nitrospir*a, we predict that this TF would regulate similar genes, such as the *frd* operon (fumarate reductase), *sdh* operon (succinate dehydrogenase), *ndh* (NADH dehydrogenase) and *ccb3* complex (cytochrome c oxidase). At least one copy of Fnr in the genomes of UW-LDO-01, *N. moscoviensis*, *N. defluvii* and N. sp. OLB-3 was located upstream of a copper-containing nitrite reductase gene (nirK), suggesting a possible mechanism that controls expression of this denitrification enzyme. The presence of multiple paralog copies of Fnr in several *Nitrospira* genomes may indicate a rigorous regulation of metabolism when these microorganisms are exposed to low levels of oxygen, an important factor affecting *Nitrospira* community compositions in nitrifying systems.[75]

Overall, this study sheds light about differences in the physiological role of NOB and comammox-like *Nitrospira*. Specifically, the comparative genomic results evidence traits associated with energy metabolism as characteristic to each of these functional groups. Furthermore, the analysis of TFs in *Nitrospira* reveals the alternative use of organic compounds, response to environmental stress, chemotaxis and anaerobic metabolism as some of the key mechanisms for the adaptive metabolism of the genus to multiple and adverse conditions. Further studies in the field should include experiments that combine omics-analysis (transcriptomics, metabolomics, and proteomics) with chemical data to confirm the ecological role and functionality of each of these functional groups and their interactions with other microorganisms.

ACKNOWLEDGMENTS

449 This work was partially supported by funding from the National Science Foundation (CBET-
450 1435661 and MCB-1518130) and the Madison Metropolitan Sewerage District. Additional
451 funding from the Chilean National Commission for Scientific and Technological Research
452 (CONICYT) as a fellowship to Pamela Camejo is also acknowledged. The U.S. Environmental
453 Protection Agency, through its Office of Research and Development, partially funded and
454 collaborated in the research described herein. Any opinions expressed in this paper are those of
455 the authors and do not necessarily reflect the views of the agency; therefore, no official
456 endorsement should be inferred. Any mention of trade names or commercial products does not
457 constitute endorsement or recommendation for use.

458 # FIGURES



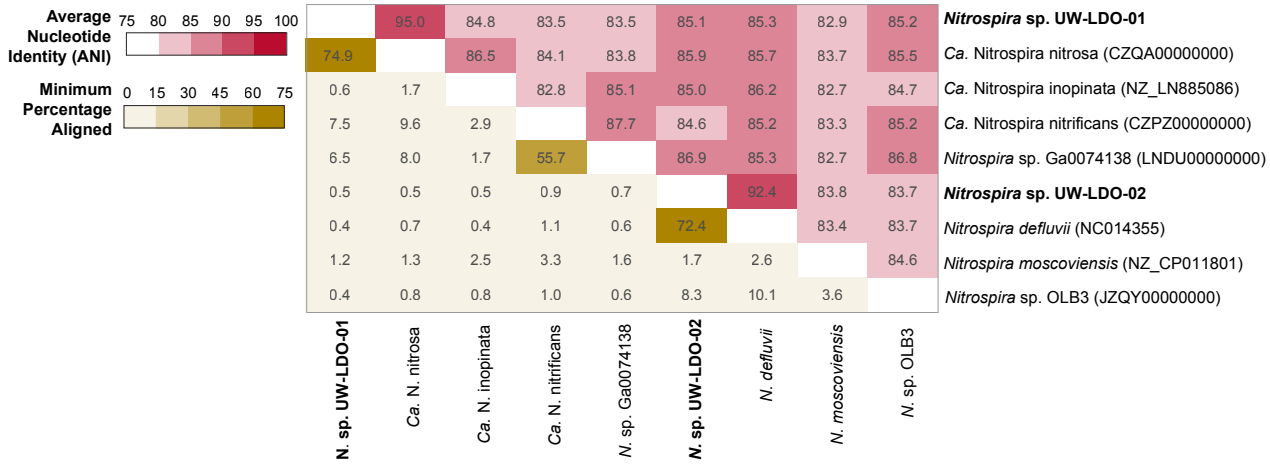
459

460 **Figure 1.** Nutrient profiles of (A) nitrogenous compounds, (B) phosphorous and acetate, and (C)

461 oxygen concentration in a regular cycle of the lab-scale SBR during Stage 1. ANA: Anaerobic,

462 AER: Micro-aerobic, SET: Settling.

463



464

465 **Figure 2.** Comparison of the genome-wide average nucleotide identity and alignment percentage,

466 of *Nitrospira*-like genomes. The heatmap shows the average nucleotide identity (red upper section

467 of matrix) and the percentage of the two genomes that aligned (yellow lower section).

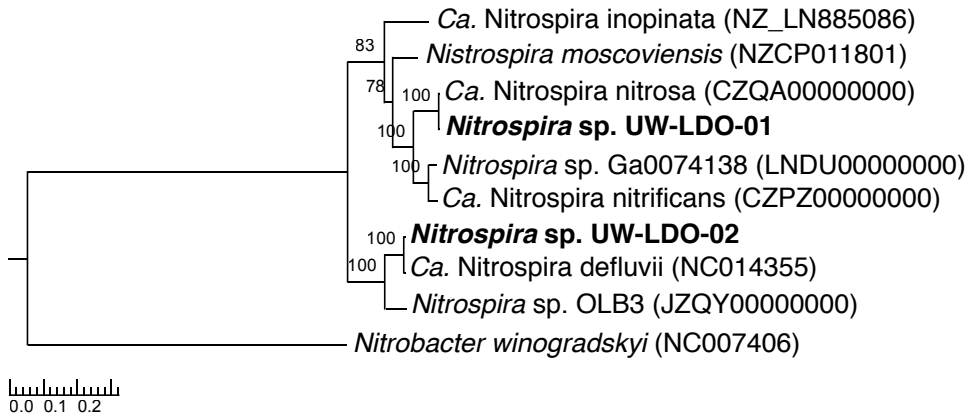


Figure 3. RAxML phylogenetic tree of a concatenated alignment of 37 marker genes (nucleotide sequence) data set with the root placed on the branch leading to *Nitrobacter winogradskyi.* The numbers at the nodes of both trees show support values derived from 100 RAxML bootstrap.

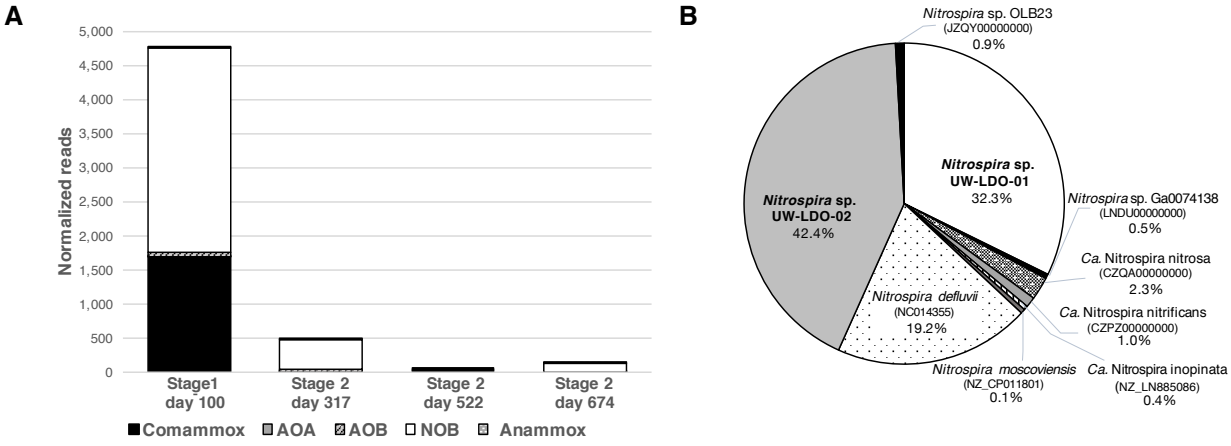


Figure 4. (A) Normalized frequency of metagenomic reads mapping to the genome of Comammox, AOB, AOA, NOB and Anammox-related organisms in samples from Stage 1 and 2 of the lab-scale SBR. (B) Relative abundance of reads mapping to genomes of *Nitrospira*-related bacteria in Stage 1 sample, including the draft genomes retrieved in this study.

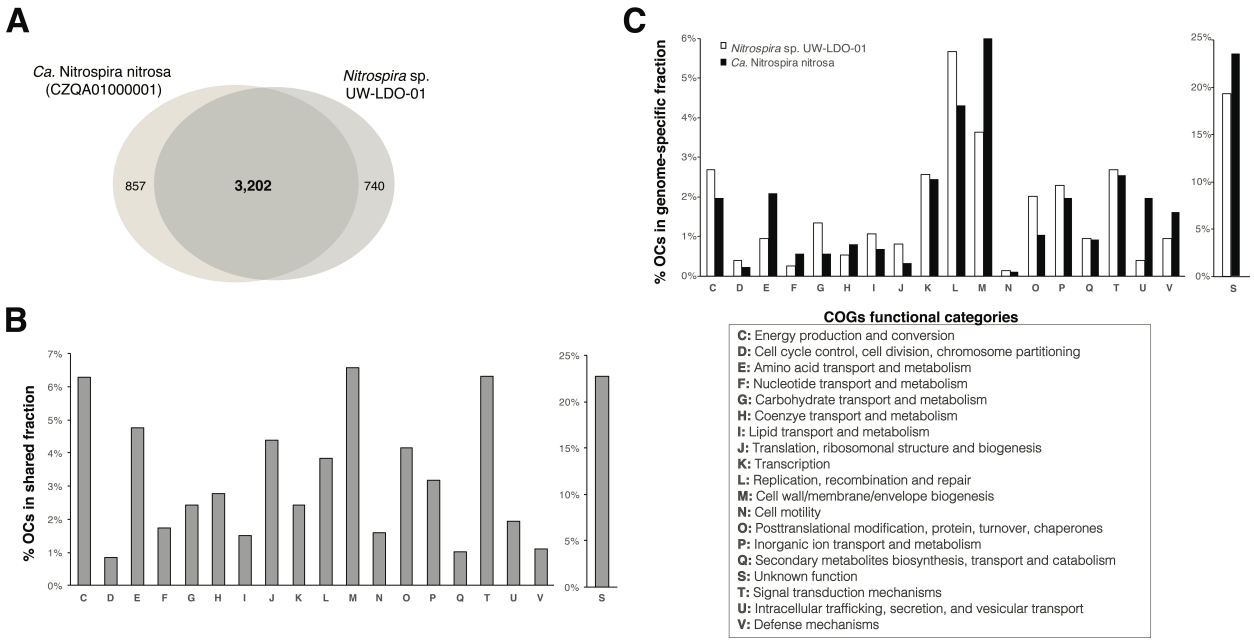


Figure 5. Genomic differences between *Nitrospira sp.* UW-LDO-01 and *Ca*. Nitrospira nitrosa. (A) Venn diagram of ortholog clusters shared between the two draft genomes; (B) distribution of COGs functional classes in the fraction of orthologs shared by the two genomes; (C) distribution of COGs functional classes in ortholog clusters found in only one of the genomes (genome-specific fraction).

TABLES

Table 1. Inventory of transcriptional regulators with implications on adaptive metabolism, from complete and draft genomes of *Nitrospira*. Grey and white rectangles represent presence and absence of each gene, respectively.

		Comammox					NOB				
Transcriptional Regulator	**Gene**	***Nitrospira* sp. UW-LDO-01**	***Ca.* Nitrospira nitrosa**	***Ca.* Nitrospira nitrificans**	***Ca.* Nitrospira inopinata**	***Nitrospira* sp. Ga0074138**	***Nitrospira* sp. UW-LDO-02**	***Nitrospira moscoviensis***	***Nitrospira defluvii***	***Nitrospira* sp. OL23**	**Function**
Formate hydrogenlyase transcriptional activator	*fhlA*										FhlA binds to formate hydrogenlyase structural genes (formate dehydrogenase and group 4 hydrogenase complex) to activate transcription of their promoters [66].
Transcriptional activator protein NhaR	*nhaR*										NhaR regulates *nhaA*, a pH-dependent sodium-proton antiporter that responds to alkaline and saline conditions [76]. It is also responsible for *osmC* induction [67], required for resistance to organic peroxides and osmotic conditions and for long-term survival in stationary phase [77, 78]. NhaR also stimulates *pga* transcription, a set of genes responsible of poly-β-1,6-N-acetyl-D-glucosamine (PGA) synthesis [68]. PGA is involved in cell-cell adhesion and attachment, which stabilizes biofilm formation [79].
Hydrogen peroxide-inducible genes activator	*oxyR*										OxyR is required for the induction of a hydrogen peroxide-inducible regulon in response to elevated levels of hydrogen peroxide [69].
Chemotaxis regulator CheZ	*cheZ*										CheZ is a component of the chemotaxis signal-transduction pathway [80]. It controls the phosphorylation of CheY, a protein involved in the cell excitation response. Absence of CheZ results in non-chemotactic cells or long stimuli response latencies, demonstrating its critical importance during response to stimuli [81].
Fumarate and nitrate reductase regulatory protein	*fnr*										Fnr is an oxygen-responsive regulator required for the expression of a number of genes involved in anaerobic metabolism [73, 82, 83], including fumarate reductase, nitrate and nitrite reductase, and cytochrome oxidase genes [84]

ASSOCIATED CONTENT

Supplementary Material. Supplementary Figures and Tables supporting the information presented in this manuscript.

REFERENCES

1. Daims, H.; Lebedeva, E. V.; Pjevac, P.; Han, P.; Herbold, C.; Albertsen, M.; Jehmlich, N.; Palatinszky, M.; Vierheilig, J.; Bulaev, A.; Kirkegaard, R. H.; von Bergen, M.; Rattei, T.; Bendinger, B.; Nielsen, P. H.; Wagner, M., Complete nitrification by Nitrospira bacteria. *Nature* **2015,** *528*, (7583), 504-9.

2. van Kessel, M. A. H. J.; Speth, D. R.; Albertsen, M.; Nielsen, P. H.; Op den Camp, H. J. M.; Kartal, B.; Jetten, M. S. M.; Lucker, S., Complete nitrification by a single microorganism. *Nature* **2015,** *528*, (7583), 555-+.

3. Palomo, A.; Jane Fowler, S.; Gulay, A.; Rasmussen, S.; Sicheritz-Ponten, T.; Smets, B. F., Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrospira spp. *ISME J* **2016,** *10*, (11), 2569-2581.

4. Pjevac, P.; Schauberger, C.; Poghosyan, L.; Herbold, C. W.; van Kessel, M. A. H. J.; Daebeler, A.; Steinberger, M.; Jetten, M. S. M.; Lücker, S.; Wagner, M.; Daims, H., Amo-targeted polymerase chain reaction primers for the specific detection and quantification of comammox Nitrospira in the environment. *bioRxiv* **2017**.

5. Okabe, S.; Satoh, H.; Watanabe, Y., In situ analysis of nitrifying biofilms as determined by in situ hybridization and the use of microelectrodes. *Appl Environ Microb* **1999,** *65*, (7), 3182-3191.

6. Daims, H.; Nielsen, J. L.; Nielsen, P. H.; Schleifer, K. H.; Wagner, M., In situ characterization of Nitrospira-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **2001,** *67*, (11), 5273-84.

7. Juretschko, S.; Timmermann, G.; Schmid, M.; Schleifer, K. H.; Pommerening-Roser, A.; Koops, H. P.; Wagner, M., Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. *Appl Environ Microbiol* **1998,** *64*, (8), 3042-51.

8. Zilles, J. L.; Peccia, J.; Kim, M. W.; Hung, C. H.; Noguera, D. R., Involvement of Rhodocyclus-related organisms in phosphorus removal in full-scale wastewater treatment plants. *Appl Environ Microb* **2002,** *68*, (6), 2763-2769.

9. Park, H. D.; Whang, L. M.; Reusser, S. R.; Noguera, D. R., Taking advantage of aerated-anoxic operation in a full-scale University of Cape Town process. *Water Environment Research* **2006,** *78*, (6), 637-642.

10. Camejo, P. Y.; Owen, B. R.; Martirano, J.; Ma, J.; Kapoor, V.; Santo Domingo, J.; McMahon, K. D.; Noguera, D. R., Candidatus Accumulibacter phosphatis clades enriched under cyclic anaerobic and microaerobic conditions simultaneously use different electron acceptors. *Water Research* **2016,** *102*, 125-137.

11. APHA In *Standard methods for the examination of water and wastewater*, Washington, DC, 2005; Andrew D. Eaton, L. S. C., Eugene W. Rice, Arnold E. Greenberg, Ed. Washington, DC, 2005.

12. Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L., IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012,** *28*, (11), 1420-1428.

544 13. Karst, S. M.; Kirkegaard, R. H.; Albertsen, M., mmgenome: a toolbox for reproducible
545 genome extraction from metagenomes. *bioRxiv* **2016**.

546 14. Albertsen, M.; Karst, S. M.; Ziegler, A. S.; Kirkegaard, R. H.; Nielsen, P. H., Back to
547 Basics--The Influence of DNA Extraction and Primer Choice on Phylogenetic Analysis of
548 Activated Sludge Communities. *PLoS One* **2015,** *10*, (7), e0132783.

549 15. Boetzer, M.; Henkel, C. V.; Jansen, H. J.; Butler, D.; Pirovano, W., Scaffolding pre-
550 assembled contigs using SSPACE. *Bioinformatics* **2011,** *27*, (4), 578-579.

551 16. Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W., CheckM:
552 assessing the quality of microbial genomes recovered from isolates, single cells, and
553 metagenomes. *Genome Res* **2015,** *25*, (7), 1043-55.

554 17. Hanson, N. W.; Konwar, K. M.; Hawley, A. K.; Altman, T.; Karp, P. D.; Hallam, S. J.,
555 Metabolic pathways for the whole community. *BMC Genomics* **2014,** *15*, 619.

556 18. Richter, M.; Rossello-Mora, R., Shifting the genomic gold standard for the prokaryotic
557 species definition. *Proceedings of the National Academy of Sciences of the United States of*
558 *America* **2009,** *106*, (45), 19126-19131.

559 19. Darling, A. E.; Jospin, G.; Lowe, E.; Matsen, F. A. t.; Bik, H. M.; Eisen, J. A., PhyloSift:
560 phylogenetic analysis of genomes and metagenomes. *PeerJ* **2014,** *2*, e243.

561 20. Stamatakis, A., RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
562 with thousands of taxa and mixed models. *Bioinformatics* **2006,** *22*, (21), 2688-90.

563 21. Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W., GenBank.
564 *Nucleic Acids Research* **2016,** *44*, (D1), D67-D72.

565 22. Wright, E. S., DECIPHER: harnessing local sequence context to improve protein
566 multiple sequence alignment. *Bmc Bioinformatics* **2015,** *16*.

567 23. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S., MEGA6: Molecular
568 Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **2013,** *30*, (12),
569 2725-9.

570 24. Muller, J.; Muller, K., TREEGRAPH: automated drawing of complex tree figures using
571 an extensible tree description format. *Molecular Ecology Notes* **2004,** *4*, (4), 786-788.

572 25. Li, L.; Stoeckert, C. J.; Roos, D. S., OrthoMCL: Identification of ortholog groups for
573 eukaryotic genomes. *Genome Research* **2003,** *13*, (9), 2178-2189.

574 26. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M. C.;
575 Rattei, T.; Mende, D. R.; Sunagawa, S.; Kuhn, M.; Jensen, L. J.; von Mering, C.; Bork, P.,
576 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for
577 eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **2016,** *44*, (D1), D286-
578 D293.

579 27. Bin, Z.; Bin, X.; Zhigang, Q.; Zhiqiang, C.; Junwen, L.; Taishi, G.; Wenci, Z.; Jingfeng,
580 W., Denitrifying capability and community dynamics of glycogen accumulating organisms
581 during sludge granulation in an anaerobic-aerobic sequencing batch reactor. *Sci Rep* **2015,** *5*,
582 12904.

583 28. Konstantinidis, K. T.; Tiedje, J. M., Genomic insights that advance the species
584 definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United*
585 *States of America* **2005,** *102*, (7), 2567-2572.

586 29. Wu, D. Y.; Jospin, G.; Eisen, J. A., Systematic Identification of Gene Families for Use
587 as "Markers" for Phylogenetic and Phylogeny-Driven Ecological Studies of Bacteria and
588 Archaea and Their Major Subgroups. *Plos One* **2013,** *8*, (10).

589 30. Keluskar, R.; Desai, A., Evaluation of hydroxylamine oxidoreductase as a functional
590 and phylogenetic marker to differentiate Nitrosomonas spp. *Journal of Basic Microbiology*
591 **2014,** *54*, (4), 261-268.

592 31. Junier, P.; Molina, V.; Dorador, C.; Hadas, O.; Kim, O. S.; Junier, T.; Witzel, K. P.;
593 Imhoff, J. F., Phylogenetic and functional marker genes to study ammonia-oxidizing
594 microorganisms (AOM) in the environment. *Applied Microbiology and Biotechnology* **2010,**
595 *85*, (3), 425-440.

596 32. Purkhold, U.; Wagner, M.; Timmermann, G.; Pommerening-Roser, A.; Koops, H. P.,
597 16S rRNA and amoA-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing
598 isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads.
599 *International Journal of Systematic and Evolutionary Microbiology* **2003,** *53*, 1485-1494.

600 33. Pester, M.; Rattei, T.; Flechl, S.; Grongroft, A.; Richter, A.; Overmann, J.; Reinhold-
601 Hurek, B.; Loy, A.; Wagner, M., amoA-based consensus phylogeny of ammonia-oxidizing
602 archaea and deep sequencing of amoA genes from soils of four different geographic regions.
603 *Environmental Microbiology* **2012,** *14*, (2), 525-539.

604 34. Fitzgerald, C. M.; Camejo, P.; Oshlag, J. Z.; Noguera, D. R., Ammonia-oxidizing
605 microbial communities in reactors with efficient nitrification at low-dissolved oxygen. *Water*
606 *Res* **2015,** *70*, 38-51.

607 35. Davey, M. E.; Duncan, M. J., Enhanced biofilm formation and loss of capsule synthesis:
608 deletion of a putative glycosyltransferase in Porphyromonas gingivalis. *J Bacteriol* **2006,** *188*,
609 (15), 5510-23.

610 36. Li, J. Y.; Wang, N., The gpsX gene encoding a glycosyltransferase is important for
611 polysaccharide production and required for full virulence in Xanthomonas citri subsp citri. *Bmc*
612 *Microbiology* **2012,** *12*.

613 37. Tao, F.; Swarup, S.; Zhang, L. H., Quorum sensing modulation of a putative
614 glycosyltransferase gene cluster essential for Xanthomonas campestris biofilm formation.
615 *Environ Microbiol* **2010,** *12*, (12), 3159-70.

616 38. Theilacker, C.; Sava, I.; Sanchez-Carballo, P.; Bao, Y.; Kropec, A.; Grohmann, E.;
617 Holst, O.; Huebner, J., Deletion of the glycosyltransferase bgsB of Enterococcus faecalis leads
618 to a complete loss of glycolipids from the cell membrane and to impaired biofilm formation.
619 *BMC Microbiol* **2011,** *11*, 67.

620 39. Wallden, K.; Rivera-Calzada, A.; Waksman, G., Type IV secretion systems: versatility
621 and diversity in function. *Cell Microbiol* **2010,** *12*, (9), 1203-12.

622 40. Klausen, M.; Heydorn, A.; Ragas, P.; Lambertsen, L.; Aaes-Jorgensen, A.; Molin, S.;
623 Tolker-Nielsen, T., Biofilm formation by Pseudomonas aeruginosa wild type, flagella and type
624 IV pili mutants. *Mol Microbiol* **2003,** *48*, (6), 1511-24.

625 41. Zechner, E. L.; Lang, S.; Schildbach, J. F., Assembly and mechanisms of bacterial type
626 IV secretion machines. *Philos Trans R Soc Lond B Biol Sci* **2012,** *367*, (1592), 1073-87.

627 42. Pascoe, B.; Meric, G.; Murray, S.; Yahara, K.; Mageiros, L.; Bowen, R.; Jones, N. H.;
628 Jeeves, R. E.; Lappin-Scott, H. M.; Asakura, H.; Sheppard, S. K., Enhanced biofilm formation

629 and multi-host transmission evolve from divergent genetic backgrounds in Campylobacter
630 jejuni. *Environ Microbiol* **2015,** *17*, (11), 4779-89.

631 43. Bronnec, V.; Turonova, H.; Bouju, A.; Cruveiller, S.; Rodrigues, R.; Demnerova, K.;
632 Tresse, O.; Haddad, N.; Zagorec, M., Adhesion, Biofilm Formation, and Genomic Features of
633 Campylobacter jejuni Bf, an Atypical Strain Able to Grow under Aerobic Conditions. *Front*
634 *Microbiol* **2016,** *7*, 1002.

635 44. Wong, E. H. J.; Ng, C. G.; Chua, E. G.; Tay, A. C. Y.; Peters, F.; Marshall, B. J.; Ho,
636 B.; Goh, K. L.; Vadivelu, J.; Loke, M. F., Comparative Genomics Revealed Multiple
637 Helicobacter pylori Genes Associated with Biofilm Formation In Vitro. *Plos One* **2016,** *11*,
638 (11).

639 45. Chipasa, K. B.; Medrzycka, K., Behavior of lipids in biological wastewater treatment
640 processes. *J Ind Microbiol Biotechnol* **2006,** *33*, (8), 635-45.

641 46. Koch, H.; Lucker, S.; Albertsen, M.; Kitzinger, K.; Herbold, C.; Spieck, E.; Nielsen, P.
642 H.; Wagner, M.; Daims, H., Expanded metabolic versatility of ubiquitous nitrite-oxidizing
643 bacteria from the genus Nitrospira. *Proc Natl Acad Sci U S A* **2015,** *112*, (36), 11371-6.

644 47. Nowka, B.; Off, S.; Daims, H.; Spieck, E., Improved isolation strategies allowed the
645 phenotypic differentiation of two Nitrospira strains from widespread phylogenetic lineages.
646 *FEMS Microbiol Ecol* **2015,** *91*, (3).

647 48. Deboer, W.; Laanbroek, H. J., Ureolytic Nitrification at Low Ph by Nitrosospira Spec.
648 *Arch Microbiol* **1989,** *152*, (2), 178-181.

649 49. Koops, H. P.; Bottcher, B.; Moller, U. C.; Pommereningroser, A.; Stehr, G.,
650 Classification of 8 New Species of Ammonia-Oxidizing Bacteria - Nitrosomonas-Communis

651 Sp-Nov, Nitrosomonas-Ureae Sp-Nov, Nitrosomonas-Aestuarii Sp-Nov, Nitrosomonas-
652 Marina Sp-Nov, Nitrosomonas-Nitrosa Sp-Nov, Nitrosomonas-Eutropha Sp-Nov,
653 Nitrosomonas-Oligotropha Sp-Nov and Nitrosomonas-Halophila Sp-Nov. *J Gen Microbiol*
654 **1991,** *137*, 1689-1699.

655 50. Koper, T. E.; El-Sheikh, A. F.; Norton, J. M.; Klotz, M. G., Urease-encoding genes in
656 ammonia-oxidizing bacteria. *Appl Environ Microb* **2004,** *70*, (4), 2342-2348.

657 51. Palatinszky, M.; Herbold, C.; Jehmlich, N.; Pogoda, M.; Han, P.; von Bergen, M.;
658 Lagkouvardos, I.; Karst, S. M.; Galushko, A.; Koch, H.; Berry, D.; Daims, H.; Wagner, M.,
659 Cyanate as an energy source for nitrifiers. *Nature* **2015,** *524*, (7563), 105-+.

660 52. Qian, M. W.; Eaton, J. W.; Wolff, S. P., Cyanate-mediated inhibition of neutrophil
661 myeloperoxidase activity. *Biochemical Journal* **1997,** *326*, 159-166.

662 53. Purcarea, C.; Ahuja, A.; Lu, T.; Kovari, L.; Guy, H. I.; Evans, D. R., Aquifex aeolicus
663 aspartate transcarbamoylase, an enzyme specialized for the efficient utilization of unstable
664 carbamoyl phosphate at elevated temperature. *Journal of Biological Chemistry* **2003,** *278*, (52),
665 52924-52934.

666 54. Ubalua, A. O., Cyanogenic Glycosides and the fate of cyanide in soil. *Australian*
667 *Journal of Crop Science* **2010,** *4*, (4), 223-237.

668 55. Widner, B.; Mulholland, M. R.; Mopper, K., Chromatographic Determination of
669 Nanomolar Cyanate Concentrations in Estuarine and Sea Waters by Precolumn Fluorescence
670 Derivatization. *Analytical Chemistry* **2013,** *85*, (14), 6661-6666.

671 56. Ushiki, N.; Fujitani, H.; Aoi, Y.; Tsuneda, S., Isolation of Nitrospira belonging to
672 Sublineage II from a Wastewater Treatment Plant. *Microbes and Environments* **2013,** *28*, (3),
673 346-353.

674 57. Gruber-Dorninger, C.; Pester, M.; Kitzinger, K.; Savio, D. F.; Loy, A.; Rattei, T.;
675 Wagner, M.; Daims, H., Functionally relevant diversity of closely related Nitrospira in
676 activated sludge. *Isme Journal* **2015,** *9*, (3), 643-655.

677 58. Koch, H.; Galushko, A.; Albertsen, M.; Schintlmeister, A.; Gruber-Dorninger, C.;
678 Lucker, S.; Pelletier, E.; Le Paslier, D.; Spieck, E.; Richter, A.; Nielsen, P. H.; Wagner, M.;
679 Daims, H., Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **2014,**
680 *345*, (6200), 1052-1054.

681 59. Silva, P. J.; de Castro, B.; Hagen, W. R., On the prosthetic groups of the NiFe
682 sulfhydrogenase from Pyrococcus furiosus: topology, structure, and temperature-dependent
683 redox chemistry. *J Biol Inorg Chem* **1999,** *4*, (3), 284-91.

684 60. Andrews, S. C.; Berks, B. C.; McClay, J.; Ambler, A.; Quail, M. A.; Golby, P.; Guest,
685 J. R., A 12-cistron Escherichia coli operon (hyf) encoding a putative proton-translocating
686 formate hydrogenlyase system. *Microbiology* **1997,** *143 (Pt 11)*, 3633-47.

687 61. Ferry, J. G., Co Dehydrogenase. *Annual Review of Microbiology* **1995,** *49*, 305-333.

688 62. Sant'Anna, F. H.; Lebedinsky, A. V.; Sokolova, T. G.; Robb, F. T.; Gonzalez, J. M.,
689 Analysis of three genomes within the thermophilic bacterial species Caldanaerobacter
690 subterraneus with a focus on carbon monoxide dehydrogenase evolution and hydrolase
691 diversity. *Bmc Genomics* **2015,** *16*.

692 63. Stover, C. K.; Pham, X. Q.; Erwin, A. L.; Mizoguchi, S. D.; Warrener, P.; Hickey, M.
693 J.; Brinkman, F. S. L.; Hufnagle, W. O.; Kowalik, D. J.; Lagrou, M.; Garber, R. L.; Goltry, L.;
694 Tolentino, E.; Westbrock-Wadman, S.; Yuan, Y.; Brody, L. L.; Coulter, S. N.; Folger, K. R.;
695 Kas, A.; Larbig, K.; Lim, R.; Smith, K.; Spencer, D.; Wong, G. K. S.; Wu, Z.; Paulsen, I. T.;
696 Reizer, J.; Saier, M. H.; Hancock, R. E. W.; Lory, S.; Olson, M. V., Complete genome sequence
697 of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* **2000,** *406*, (6799), 959-
698 964.

699 64. Gioia, J.; Qin, X.; Jiang, H. Y.; Clinkenbeard, K.; Lo, R.; Liu, Y. M.; Fox, G. E.;
700 Yerrapragada, S.; McLeod, M. P.; McNeill, T. Z.; Hemphill, L.; Sodergren, E.; Wang, Q. Y.;
701 Muzny, D. M.; Homsi, F. J.; Weinstock, G. A.; Highlander, S. K., The genome sequence of
702 Mannheimia haemolytica A1: Insights into virulence, natural competence, and Pasteurellaceae
703 phylogeny. *Journal of Bacteriology* **2006,** *188*, (20), 7257-7266.

704 65. Brinkrolf, K.; Brune, I.; Tauch, A., The transcriptional regulatory network of the amino
705 acid producer Corynebacterium glutamicum. *J Biotechnol* **2007,** *129*, (2), 191-211.

706 66. Schlensog, V.; Lutz, S.; Bock, A., Purification and DNA-binding properties of FHLA,
707 the transcriptional activator of the formate hydrogenlyase system from Escherichia coli. *J Biol*
708 *Chem* **1994,** *269*, (30), 19590-6.

709 67. Toesca, I.; Perard, C.; Bouvier, J.; Gutierrez, C.; Conter, A., The transcriptional
710 activator NhaR is responsible for the osmotic induction of osmC(p1), a promoter of the stress-
711 inducible gene osmC in Escherichia coli. *Microbiology-Sgm* **2001,** *147*, 2795-2803.

712 68. Goller, C.; Wang, X.; Itoh, Y.; Romeo, T., The cation-responsive protein NhaR of
713 Escherichia coli activates pgaABCD transcription, required for production of the biofilm
714 adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *J Bacteriol* **2006,** *188*, (23), 8022-32.

715 69. Storz, G.; Tartaglia, L. A., OxyR: a regulator of antioxidant genes. *J Nutr* **1992,** *122*,
716 (3 Suppl), 627-30.

717 70. Bueno, E.; Mesa, S.; Bedmar, E. J.; Richardson, D. J.; Delgado, M. J., Bacterial
718 adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxid*
719 *Redox Signal* **2012,** *16*, (8), 819-52.

720 71. Green, J.; Crack, J. C.; Thomson, A. J.; LeBrun, N. E., Bacterial sensors of oxygen.
721 *Curr Opin Microbiol* **2009,** *12*, (2), 145-51.

722 72. Green, J.; Scott, C.; Guest, J. R., Functional versatility in the CRP-FNR superfamily of
723 transcription factors: FNR and FLP. *Adv Microb Physiol* **2001,** *44*, 1-34.

724 73. Reents, H.; Munch, R.; Dammeyer, T.; Jahn, D.; Hartig, E., The Fnr regulon of Bacillus
725 subtilis. *J Bacteriol* **2006,** *188*, (3), 1103-12.

726 74. Salmon, K.; Hung, S. P.; Mekjian, K.; Baldi, P.; Hatfield, G. W.; Gunsalus, R. P.,
727 Global gene expression profiling in Escherichia coli K12. The effects of oxygen availability
728 and FNR. *J Biol Chem* **2003,** *278*, (32), 29837-55.

729 75. Park, H. D.; Noguera, D. R., Nitrospira community composition in nitrifying reactors
730 operated with two different dissolved oxygen levels. *J Microbiol Biotechnol* **2008,** *18*, (8),
731 1470-4.

732 76. Padan, E.; Maisler, N.; Taglicht, D.; Karpel, R.; Schuldiner, S., Deletion of ant in
733 Escherichia coli reveals its function in adaptation to high salinity and an alternative Na+/H+
734 antiporter system(s). *J Biol Chem* **1989,** *264*, (34), 20297-302.

735 77. Conter, A.; Gangneux, C.; Suzanne, M.; Gutierrez, C., Survival of Escherichia coli
736 during long-term starvation: effects of aeration, NaCl, and the rpoS and osmC gene products.
737 *Research in Microbiology* **2001,** *152*, (1), 17-26.

738 78. Mongkolsuk, S.; Praituan, W.; Loprasert, S.; Fuangthong, M.; Chamnongpol, S.,
739 Identification and characterization of a new organic hydroperoxide resistance (ohr) gene with
740 a novel pattern of oxidative stress regulation from Xanthomonas campestris pv. phaseoli. *J*
741 *Bacteriol* **1998,** *180*, (10), 2636-43.

742 79. Itoh, Y.; Wang, X.; Hinnebusch, B. J.; Preston, J. F., 3rd; Romeo, T., Depolymerization
743 of beta-1,6-N-acetyl-D-glucosamine disrupts the integrity of diverse bacterial biofilms. *J*
744 *Bacteriol* **2005,** *187*, (1), 382-7.

745 80. Bren, A.; Welch, M.; Blat, Y.; Eisenbach, M., Signal termination in bacterial
746 chemotaxis: CheZ mediates dephosphorylation of free rather than switch-bound CheY. *Proc*
747 *Natl Acad Sci U S A* **1996,** *93*, (19), 10090-3.

748 81. Khan, S.; Amoyaw, K.; Spudich, J. L.; Reid, G. P.; Trentham, D. R., Bacterial
749 chemoreceptor signaling probed by flash photorelease of a caged serine. *Biophys J* **1992,** *62*,
750 (1), 67-8.

751 82. Unden, G.; Schirawski, J., The oxygen-responsive transcriptional regulator FNR of
752 Escherichia coli: the search for signals and reactions. *Mol Microbiol* **1997,** *25*, (2), 205-10.

753 83. Spiro, S.; Guest, J. R., FNR and its role in oxygen-regulated gene expression in
754 Escherichia coli. *FEMS Microbiol Rev* **1990,** *6*, (4), 399-428.

755 84. Li, B.; Wing, H.; Lee, D.; Wu, H. C.; Busby, S., Transcription activation by Escherichia
756 coli FNR protein: similarities to, and differences from, the CRP paradigm. *Nucleic Acids Res*
757 **1998,** *26*, (9), 2075-81.
758