Unexpected versatility in the metabolism and ecophysiology of globally relevant nitrite-oxidizing *Nitrotoga* bacteria

Andrew M. Boddicker[1] and Annika C. Mosier[1†]

[1]Department of Integrative Biology, University of Colorado Denver, Campus Box 171, Denver, CO, USA

[†]To whom correspondence should be addressed. E-mail: annika.mosier@ucdenver.edu

Keywords: *Nitrotoga*, nitrite-oxidizing bacteria, freshwater nitrification

Running title: *Nitrotoga* are prevalent and versatile nitrite oxidizers

Competing Interests: The authors declare no competing interests. Portions of this manuscript were previously published as a part of University of Colorado Denver Master's thesis submission (AB, 2017). Funding was provided by the University of Colorado, Denver and the City and County of Denver.

ABSTRACT

Nitrite-oxidizing bacteria (NOB) play a critical role in the mitigation of nitrogen pollution from freshwater systems by metabolizing nitrite to nitrate, which is removed via assimilation, denitrification, or anammox. Recent studies revealed that NOB are phylogenetically and metabolically diverse, yet most of our knowledge of NOB comes from only a few cultured representatives. Using enrichment methods and genomic sequencing, we identified four novel *Candidatus* Nitrotoga NOB species from freshwater sediments and water column samples in Colorado, USA. Genome assembly revealed highly conserved 16S rRNA gene sequences, but a surprisingly broad metabolic potential including genes for nitrogen, sulfur, hydrogen, and organic carbon metabolism. Genomic predictions suggest that *Nitrotoga* can metabolize in low oxygen or anaerobic conditions, which may support a previously unrecognized environmental niche. An array of antibiotic and metal resistance genes likely allows *Nitrotoga* to withstand environmental pressures in impacted systems. Phylogenetic analyses reveal a deeply divergent nitrite oxidoreductase alpha subunit (NxrA) not represented in any other NOB, suggesting a novel evolutionary trajectory for *Nitrotoga*. *Nitrotoga*-like 16S rRNA gene sequences were prevalent in globally distributed environments. This work considerably expands our knowledge of the *Nitrotoga* genus and improves our understanding of their role in the global nitrogen cycle.

INTRODUCTION

Increasing anthropogenic sources of nitrogen have led to environmental risks including eutrophication, increased greenhouse gas emissions, and acidification. Nitrite-oxidizing bacteria (NOB) play a critical role in mitigating the harmful effects of nitrogen pollution by linking nitrogen sources to nitrogen removal. Specifically, nitrite pools from natural (e.g., ammonia oxidation) or anthropogenic (e.g., fertilizer) sources are converted into nitrate, which is assimilated or removed as inert nitrogen gas via denitrification or anammox pathways. Thus, understanding the diversity, metabolism, and environmental limits of NOB is important for controlling and managing elevated nitrogen in contaminated ecosystems.

Despite the environmental relevance of NOB, the group is understudied in part due to slow growth, taking as long as 12 years for isolation (Lebedeva, *et al*. 2008). Assiduous cultivation efforts (Alawi *et al*., 2007; Daims *et al*., 2015; Sorokin *et al*., 2012; van Kessel *et al*., 2015; Watson *et al*., 1986; Watson & Waterbury, 1971), as well as single-cell (Ngugi *et al*., 2015) and metagenomic (Pinto *et al*., 2015) sequencing studies are beginning to illuminate the diversity of nitrite oxidizers. The known NOB belong to four phyla and seven different genera, three of which were discovered within the last decade: *Candidatus* Nitrotoga, *Nitrolancea*, and *Candidatus* Nitromaritima (Alawi *et al*., 2007; Ngugi *et al*., 2015; Sorokin *et al*., 2012). NOB are physiologically versatile, utilizing nitrite oxidation, complete ammonia oxidation (comammox within the *Nitrospira*) (Daims *et al*., 2015; van Kessel *et al*., 2015), as well as organic carbon, hydrogen, and sulfur metabolisms to drive growth (Koch *et al*., 2016; Daims *et al*., 2016; Füssel *et al*., 2017).

Nitrotoga is an understudied genus of NOB that may play a critical role in nitrogen removal in engineered and natural environments (Lücker *et al*., 2015). Recent 16S rRNA-based molecular surveys have identified *Nitrotoga*-like sequences in a surprisingly wide range of habitats, including: glacial soils (Pradhan *et al*., 2010; Sattin *et al*., 2009; Schmidt *et al*., 2009; Srinivas *et al*., 2011), an underground cave (Chen *et al*., 2009), a freshwater seep (Roden *et al*., 2012), drinking water (Kinnunen *et al*., 2017; White, DeBry, & Lytle, 2012), a subglacial Antarctic lake (Christner *et al*., 2014), Yellow Sea seawater (Na *et al*., 2011), salt marsh sediments (Martiny *et al*., 2011), rivers (Fan *et al*., 2016; Li *et al*., 2011), and various wastewater treatment plants (WWTPs) or sequence batch reactors (Bereschenko *et al*., 2010; Figdore, Stensel, & Winkler, 2017; Karkman *et al*., 2011; Lücker *et al*., 2015; Saunders *et al*., 2015; Yang, Chen, & Li, 2016; Ziegler *et al*., 2016). The relative abundance of *Nitrotoga*-like sequences in several WWTPs and a subglacial lake were as high as 2-13% of the total bacterial community, and were occasionally the only observed NOB (Christner *et al*., 2014; Lücker *et al*., 2015; Saunders *et al*., 2015). The distribution and relative abundance of *Nitrotoga*-like sequences suggest these organisms likely play a critical nitrogen cycling role in diverse environments; however, few studies have attempted to characterize their diversity or physiology.

Only four *Nitrotoga* cultures have been reported to date, with no confirmed isolates and no genome sequences are available within the genus. *Candidatus* Nitrotoga arctica was enriched from permafrost (Alawi *et al*., 2007); *Candidatus* Nitrotoga sp. HAM-1 from activated sludge (Alawi *et al*., 2009); *Candidatus* Nitrotoga sp. HW29 from an aquaculture system (Hüpeden *et al*., 2016); and *Candidatus* Nitrotoga sp. AM1 from coastal sand (Ishii *et al*., 2017). All *Nitrotoga* were enriched at low temperatures (4-17°C), however temperature optima were slightly higher (13-22°C) (Alawi *et al.*, 2007, 2009; Hüpeden *et al.*, 2016; Ishii *et al.*, 2017). *Ca.* Nitrotoga arctica has an intermediate nitrite affinity and is adapted to low nitrite concentrations (0.3 mM) compared to other NOB (Nowka *et al.*, 2015; Kits *et al.*, 2017).

Here, we describe near-complete draft genome sequences of four novel *Nitrotoga* species enriched from river water column and sediment samples. Each organism contained genomic capabilities for diverse metabolisms, which could support their growth in a wide range of habitats. This study represents the first reported cultivation of *Nitrotoga* from natural, freshwater systems and the first genome profiling from within the genus. These findings extend our understanding of freshwater nitrite oxidation and form the basis for further experimental work aimed at testing genomic predictions in culture and in the environment.

METHODS

Culture inoculation and growth

Two surface sediment samples were collected from the urban-impacted Cherry Creek in downtown Denver, CO (samples MKT and LAW). Two water column samples were collected from two agriculturally-impacted rivers near Greeley, CO (about 100 km North of Denver, CO) (samples CP45 from the Cache la Poudre River and SPKER from the South Platte River). Sediment and water column samples were inoculated into Freshwater Nitrite Oxidizer Medium (FNOM) with 0.3 mM nitrite and incubated at room temperature in the dark (Supplemental Note). Enrichment cultures were transferred to new media approximately every two weeks. Enrichment was enhanced by serial dilution, rapid transfers at the beginning of nitrite consumption, and low volume transfers (as low as 0.1% inoculum vol./vol.). Nitrite consumption was regularly monitored in the cultures using a Griess nitrite color reagent (Griess-Romijn van Eck, 1966). Nitrite oxidation rates were determined in triplicate for each enrichment culture (Supplemental Note).

DNA extraction and sequencing

DNA was extracted from each culture at mid- to late-phase of exponential nitrite oxidation (Supplemental Note). Extracted DNA was sheared using a Covaris S220 (Covaris, Woburn, MA) and libraries were prepped with an insert size of 400 bp using an Ovation Ultralow System V2 (No. 0344) kit (Nugen, San Carlos, CA) by the University of Colorado Anschutz Medical Campus Genomics Core. DNA was sequenced on an Illumina HiSeq 2500 using V4 chemistry (Illumina, San Diego, CA) with 125 bp paired end reads.

Metagenome and *Nitrotoga* genome assembly and annotation

Metagenomes from each enrichment culture were assembled with quality filtered and trimmed reads using MEGAHIT (Li *et al.*, 2014) and contigs were binned using MetaBAT (Kang *et al.*, 2015). *Nitrotoga* genomes were assembled iteratively with SPAdes v3.9.0 (Bankevich *et al.*, 2012) (Supplemental Note). Genome bin completeness and contamination estimates were calculated using the CheckM v1.0.11 lineage workflow (Parks *et al.*, 2015) (Supplemental Note).

Final assemblies were filtered to remove contigs <2 kb, all of which had very low and uneven coverage estimates. The *Nitrotoga* genomes were aligned and contigs reordered with progressiveMauve (Darling *et al.*, 2010), using the *Nitrotoga* species from the MKT culture as a reference due to its long contig length and simple assembly graph. Genomes were submitted to the DOE-JGI Microbial Genome Annotation Pipeline (MGAP) (Huntemann *et al.*, 2015) for final contig trimming of ambiguous and low-complexity sequences. Gene annotations were evaluated based on results from MGAP including COG, KEGG, Pfam, and TIGRfam assignments, InterPro Scan, IMG term assignments, and final protein product assignments

(Huntemann *et al.*, 2015 and references within), as well as a BLASTP search of all predicted CDS against the nr database and annotation with KASS (Moriya *et al.*, 2007). RNAs were detected via MGAP with CRT, pilercr, tRNAscan, hmmsearch, BLASTN, and cmsearch (Huntemann *et al.*, 2015 and references within).

Nitrotoga comparative genomics

The Anvi'o pangenome pipeline (Eren *et al.*, 2015) was used to cluster (mcl=10) coding sequences from each *Nitrotoga* genome in order to establish a 'core genome' of genes shared by all four *Nitrotoga* spp., and genes unique to each genome or shared by two or three *Nitrotoga* spp. Average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated using the online enveomics tools (Rodriguez-R and Konstantinidis, 2016).

Nitrotoga phylogenetic analyses

Near full-length *Nitrotoga*-like 16S rRNA gene sequences in the NCBI nt database were identified by a BLASTN search. Sequences were retained if they were ≥1300 bp in length and had ≥95% identity to any of the cultivated *Nitrotoga* 16S rRNA gene sequences from this study. 16S rRNA gene sequences from the BLASTN search were aligned with MAFFT (Katoh and Standley, 2013) and manually trimmed. A maximum likelihood tree was generated using RAxML (version 8.2.9) with 100 rapid bootstraps and the GTRGAMMA model of nucleotide substitution. A single monophyletic clade was extracted from the resulting phylogenetic tree, which included all known *Nitrotoga* 16S rRNA gene sequences. Neighboring clades included members of different genera. Selected outgroup sequences were added to the extracted sequences and then were realigned and trimmed, and a maximum likelihood tree was generated with 1000 rapid bootstraps.

Amino acid sequences (≥833 amino acids in length) of Type II DMSO reductase family enzymes were compiled based on previous research (Lücker *et al.*, 2010, 2013; Ngugi *et al.*, 2015) and a BLASTP search of the NCBI nr database. Sequences were aligned using MAFFT (Katoh and Standley, 2013) and manually trimmed. A maximum likelihood tree was built using RAxML with 1000 rapid bootstraps and the LG likelihood model of amino acid substitution (based on PROTGAMMAAUTO selection). All trees were visualized and annotated in iTOL (Letunic and Bork, 2016).

Distribution of _Nitrotoga_-like 16S rRNA gene sequences in the environment

Full-length 16S rRNA gene sequences from each *Nitrotoga* culture were submitted to the IMNGS online server (Lagkouvardos *et al.*, 2016) for searches against all 183,153 16S rRNA gene amplicon runs from the NCBI Sequence Read Archive (SRA) (March 2018 release). A minimum identity threshold of 97% was chosen to represent *Nitrotoga*-like sequences at the genus level due to phylogenetic resolution of 16S rRNA gene sequences (see results). Samples were removed if they did not have at least 100 reads associated with *Nitrotoga*-like operational taxonomic units (OTUs). Data was summarized by SRA annotated environment categories. In some cases, SRA annotated categories were merged together: "aquatic", "freshwater", and "pond" were merged into "freshwater"; "freshwater_sediment" and "sediment" were merged into "sediment"; "biofilm" and "microbial_mat" were merged into "biofilm"; "soil", "terrestrial", and "peat" were combined into "soil"; and "plant", "rhizosphere", and "root" were merged into "plant-associated".

Relative abundance of *Nitrotoga* in creek sediments and water column

Water column and sediment samples were collected from 18 sites along Bear Creek, 18 sites along Cherry Creek, and four sites along the South Platte River at the confluences of Bear Creek and Cherry Creek in Denver, Colorado, USA in Fall 2016 (Supplemental Note). DNA extracts were sent to the University of Illinois Roy J. Carver Biotechnology Center, Urbana, Illinois for total 16S rRNA amplicon sequencing using the 515F-Y and 926R primers (Parada *et al.*, 2015; Quince *et al.*, 2011). Libraries were prepared with the Fluidigm 48.48 Access Array IFC platform (Fluidigm Corporation, South San Francisco, CA) as previously described (Ramanathan *et al.*, 2017), and sequenced on an Illumina HiSeq with Rapid 250 bp paired-end reads (Illumina, San Diego, CA). Filtered reads were clustered into OTUs at 97% sequence identity and taxonomy was assigned with a BLASTN search against the SILVA 16S rRNA gene database (release 128) (Quast *et al.*, 2013).

RESULTS AND DISCUSSION

Enrichment and nitrite oxidation

The four nitrite-oxidizing cultures were enriched for 17 (CP45 and SPKER) or 20 (LAW and MKT) months via serial dilution and rapid transfer. Nitrite oxidation rates (calculated across three measurements during logarithmic nitrite oxidation) averaged 135.6 ± 21.5 μM NO_2^-/day (Figure 1), which was similar to previously reported data for *Ca.* Nitrotoga arctica (Nowka *et al.*, 2015). 16S rRNA gene sequence analyses and PCR with NOB specific primers revealed that each culture contained only one NOB related to the *Candidatus* Nitrotoga genus (Betaproteobacteria class; Nitrosomonadales order; Gallionellaceae family).

We propose the following names for the enriched *Nitrotoga* species: *Candidatus* Nitrotoga mira MKT (n, nominative adjective, *mirus*, surprising) for the species enriched from Cherry Creek sediment and the unexpected enrichment of freshwater *Nitrotoga*; *Candidatus* Nitrotoga auraria LAW (f, nominative, noun, *aurarius*, of gold) for the species enriched from Cherry Creek sediment located near the historic Auraria settlement; *Candidatus* Nitrotoga amnis CP45 (m, genitive, noun, *amnis*, river) for the species enriched from Cache la Poudre River water; and *Candidatus* Nitrotoga coloradensis SPKER (f, adjective, Colorado, state central US) for the species enriched from water in the South Platte River, which flows through northeastern Colorado. Strain names represent sampling sites specific to each culture.

Metagenome assembly and binning

Metagenomic sequencing of the enrichment cultures showed that each culture contained 8-32 genome bins (CP45: 8; MKT: 14; LAW: 23; SPKER: 32), likely corresponding to a similar number of species in each enrichment culture. Attempted co-assemblies of all four metagenomes were not successful as each enrichment culture had different species compositions. Each metagenome assembly contained only one predicted NOB related to *Nitrotoga* spp, along with other Proteobacteria including *Pseudomonas* spp, *Methylotenera* spp, and several members of the Comamonadaceae family (based on EMIRGE-assembled 16S rRNA genes (Miller *et al.*, 2011) and the CheckM lineage workflow). One genome bin from the SPKER metagenome likely belonged to a protozoan in the Neobodonida order that may prey on bacteria in culture.

Nitrotoga genome assembly

The *Nitrotoga* genome bins from each culture had very high average coverage (201-398X). *Nitrotoga* genome bins were predicted to be 99.8% complete with ≤0.3% contamination by CheckM (after manual assignment of some marker genes; Supplemental Note; Supplementary Table S1). These high-quality draft sequences are the first known *Nitrotoga* genomes. The closest relatives with genome sequences were *Sideroxydans lithotrophicus* ES-1, *Gallionella capsiferriformans* ES-2, and *Ca.* Gallionella acididurans ShG14-8.

Nitrotoga genomes ranged in size from 2.707-2.982 Mbp, with 23-59 contigs and GC content between 47.5% and 48.8% (Supplementary Table S1). The number of coding sequences ranged from 2,574-2,858 with 36-39 tRNAs encoding all twenty amino acids. ANI values between the four *Nitrotoga* genomes ranged from 85.9%-93.4% (Table 1), while AAI values had a similar range of 87.5%-94.6%. These ANI values were indicative of each enrichment containing a novel *Nitrotoga* species, as they fell below the assumed 95% ANI threshold that separates most bacterial species (Caro-Quintero & Konstantinidis, 2012; Jain *et al*., 2017; Konstantinidis, Rosselló-móra, & Amann, 2017; Rodriguez-R & Konstantinidis, 2014). AAI values indicated these organisms were still highly conserved and likely shared many of the same traits. The SPKER *Nitrotoga* genome was considerably more divergent than the CP45, MKT, or LAW genomes.

Anvi'o pangenome analysis showed that coding sequences from all four genomes (10,666 in total) grouped into 4,001 protein clusters (PCs) (Figure 2). The core genome was represented by 1,803 PCs found in all four *Nitrotoga* genomes (45.1% of total). Each individual genome (e.g., MKT only) contained 293-625 unique PCs, while 566 PCs were shared among two or three genomes. Of the 2,198 accessory genome PCs, 1,041 were annotated as hypothetical proteins.

Despite genomic level differences, the *Nitrotoga* 16S rRNA gene sequences were highly conserved (Figure 3; Supplementary Figure S1). *Nitrotoga* 16S rRNA genes were present at double coverage in all *Nitrotoga* genomes, indicating a gene duplication. The 16S rRNA gene from the SPKER genome had three single nucleotide variants (SNVs) present in ~50% of mapped reads, likely indicating small differences between the duplicate copies. A consensus sequence was used here, as the variants could not be isolated with paired reads. Among the four *Nitrotoga* 16S rRNA gene sequences assembled in this study, the most divergent are 99.4% identical across a 1,544 bp alignment (9 total mismatches). Pairwise comparisons of all 8 enriched *Nitrotoga* 16S rRNA gene sequences (4 previous studies, 4 from this study) average 99.5% identical across near full-length genes (Figure 3b) and a comparison of all 60 *Nitrotoga*-like sequences in Figure 3, comprising the known *Nitrotoga* genus, averaged 98.7% (Supplemental Figure 1). The highly conserved nature of *Nitrotoga* 16S rRNA gene sequences does not likely represent genome-level conservation across the genus, based on 16S rRNA:genome comparisons in the present study.

Nitrotoga nitrogen metabolism

Nitrite oxidation

Nitrite oxidoreductase (NXR) is a heterotrimeric enzyme (NxrABC) that oxidizes nitrite to nitrate, liberating two electrons from a water molecule (Supplemental Note). NXR is a member of the Type II DMSO reductase family of molybdenum enzymes. NXR is bound to the cell membrane and can be classified into at least two distinct phylogenetic and functional groups based on orientation towards the cytoplasm or periplasm (Lücker *et al.*, 2010). Oxidation kinetics studies have associated NXR orientation with ecological niche formation, as bacteria with

cytoplasmic-facing NXR (*Nitrobacter*, *Nitrococcus*, and *Nitrolancea*) typically dominate in relatively high nitrite environments over bacteria with periplasmic-facing NXR (*Nitrospira*, *Nitrospina*, and *Candidatus* Nitromaritima) (Kits *et al*., 2017; Koch *et al*., 2015; Lücker *et al*., 2010; Lücker, *et al*. 2013; Nowka *et al*., 2015; Sorokin *et al*., 2012; Spieck, *et al*. 1996; Spieck, *et al*. 1998; Starkenburg *et al*., 2006). Periplasmic-facing NXR have the energetic benefit of H^+ release contributing to the proton motive force, while nitrite must be pumped into the cell for cytoplasmic-facing NXR.

Nitrotoga nxr genes were found on single contigs (6.7-9.5kb) within each genome forming an *nxrABC* operon, with an additional *nxrD* chaperone (Supplemental Note). Based on the number of neighboring contigs in the assembly and read coverage estimates, the *nxr* genes in the CP45, LAW, and MKT genomes are thought to be duplicated. The SPKER *nxr* operon is likely present three times throughout the genome based on the same criteria.

Nitrotoga NxrA amino acid sequences were divergent from all other members of the Type II DMSO reductase enzyme family, and the closest relatives were putative archaeal nitrate reductase (NarG) proteins (Figure 4a). The deeply branching *Nitrotoga* NxrA amino acid sequences may represent a fourth evolutionary development of nitrite oxidation, separate from the cytoplasmic-facing, periplasmic-facing, and phototrophic *Thiocapsa* nitrite oxidizers (Hemp *et al.*, 2016; Lücker *et al.*, 2010). Interestingly, some contigs surrounding the *Nitrotoga nxr* operons contained putative transposase and integrase genes, possibly suggesting that the *Nitrotoga nxr* genes were horizontally transferred similar to the theory of periplasmic-facing *nxr* gene development (Lücker *et al.*, 2010). The *Nitrotoga* genus also represents the only known Betaproteobacterial lineage of nitrite-oxidation.

Nitrotoga NxrA from the CP45, LAW, and MKT genomes had a predicted twin-arginine translocation (Tat) signal peptide on the N-terminus (Figure 4b), which supports the excretion of this enzyme into the periplasm (Sargent, 2007). The SPKER genome NxrA was lacking a signal peptide, although an alignment showed that the first 14 amino acids may be missing, as the entire peptide was otherwise aligned without gaps. The gene was located at the end of a contig and was missing a start codon.

Conserved residues were previously predicted to play a role in binding a molybdopterin cofactor, a [4Fe-4S] cluster, and the nitrite/nitrate substrate in NxrA and nitrate reductase NarG (Martinez-espinosa *et al.*, 2007; Lücker *et al.*, 2010). All *Nitrotoga* NxrA shared the conserved residues for the molybdopterin, iron-sulfur cluster, and four of the five nitrite/nitrate binding residues. However, an analysis of the fifth residue revealed a surprising lack of conservation among all NOB with at least five different residues predicted in known NxrA sequences, including variations within *nxrA* copies from a single genome (e.g., *Nitrospira nitrosa* has two *nxrA* copies, one with asparagine and one with glycine in this same position) (Figure 4b). This residue may not play a critical role in nitrite/nitrate binding, or may play a role in the variable nitrite oxidation kinetics observed previously (Nowka *et al.*, 2015; Kits *et al.*, 2017).

Nitrotoga NxrB were lacking a signal peptide, but may be excreted into the periplasm via a 'hitch-hiker' method as described previously (Lücker *et al.*, 2010; Martinez-espinosa *et al.*, 2007). All NxrB had coordinating residues for three [4Fe-4S] and one [3Fe-4S] cluster for electron conductance (Supplemental Figure 2).

One *Nitrotoga* NxrC predicted protein sequence was found in each genome with an N-terminal signal peptide for excretion. Signal peptides were previously observed in *Nitrospina gracilis* and *Candidatus* Nitromaritima NxrC (Lücker, *et al*. 2013; Ngugi, *et al*. 2015), however the predicted signal peptide cleavage site was located in the middle of a predicted

transmembrane region. Biochemical studies confirmed the membrane-association of the NXR holoenzyme in *Nitrospina* and *Nitrospira* (Bartosch *et al*., 1999), indicating that NxrC were not fully translocated into the periplasm (Lücker *et al.*, 2010). *Nitrotoga* NxrC do not have predicted transmembrane regions, and are more similarly related to the soluble periplasmic gamma subunit of ethylbenzene dehydrogenase, which likely interact with cytochrome *c* proteins to shuttle electrons to the membrane (Kloer, *et al*. 2006); similar observations were made in the analysis of *Nitrospina gracilis* and *Candidatus* Nitromaritima genomes (Lücker, *et al*. 2013; Ngugi, *et al*. 2015).

Overall, sequence and phylogenetic analyses indicated that *Nitrotoga* possess a form of NXR that is divergent from known NOB. Phylogenetically, the alpha subunit was deeply branched near putative bacterial and archaeal NarG, but maintains all necessary residues for nitrite oxidation. The NxrA and NxrC subunits had signal peptides for excretion to the periplasm, but all subunits were lacking transmembrane domains for anchoring in the cytoplasmic membrane. This may suggest that *Nitrotoga* have a soluble NXR periplasmic holoenzyme, which has never been observed in NOB.

Dissimilatory and assimilatory nitrogen metabolism

All *Nitrotoga* genomes had genes for a NirK dissimilatory nitrite reductase for the reduction of nitrite to nitric oxide. *nirK* genes have been found in all other NOB genomes except *Nitrolancea hollandica* (Lücker *et al.*, 2013; Sorokin *et al.*, 2012), but their ultimate role is still unclear. CP45, LAW, and MKT genomes encoded a nitric oxide dioxygenase (*hmp*), which catalyzes the conversion of nitric oxide to nitrate, and is evolutionarily related to an O_2-binding protein similar to hemoglobin (Gardner *et al.*, 1998). The role of *nirK* and *hmp* in *Nitrotoga* is unknown.

The *Nitrotoga* genomes encoded genes for transport of nitrite/nitrate (*narK*), formate/nitrite, and ammonium (*amtB*). The predicted periplasmic orientation of *Nitrotoga* NXR excludes the need for nitrite import into the cytosol, however genes for an assimilatory NirBD nitrite reductase and cytochrome c551/c552 to catalyze the reduction of nitrite to ammonia were found. *Nitrotoga* may also assimilate ammonia released during cyanide detoxification (Supplemental Note). A *Nitrotoga* enrichment culture from coastal sands grew faster when ammonium was added to the culture medium (Ishii *et al.*, 2017), likely due to the reduced need for assimilatory nitrite reduction.

***Nitrotoga* energy metabolism and reverse electron flow**

Nitrogen energetics

A complete electron transport chain was present in the *Nitrotoga* genomes (Supplemental Figure 3). After electrons are passed from NXR to cytochrome *c*, they are transferred to oxygen via a terminal oxidase (Complex IV). All *Nitrotoga* genomes contained genes for a *cbb*$_3$-type cytochrome *c* oxidase, a member of the C-class heme-copper oxidases with an exceptionally high affinity for oxygen (Morris and Schmidt, 2013). These genes did not form a distinct operon, however there were no other candidate terminal oxidase genes in most *Nitrotoga* genomes (see below). Organisms possessing *cbb*$_3$-type oxidases, including the NOB *Nitrospina gracilis* and the phototrophic nitrite-oxidizer *Thiocapsa* KS1, are likely capable of growth in microoxic environments (Han *et al.*, 2011; Lücker *et al.*, 2013; Hemp *et al.*, 2016). For instance, *Nitrospina*-like bacteria have been found to play a crucial role in carbon fixation in marine oxygen minimum zones, due in part to their *cbb3*-type terminal oxidases (Füssel *et al.*, 2012,

2017; Pachiadaki *et al.*, 2017). The possession of a *cbb*$_3$-type terminal oxidase indicates that *Nitrotoga* species likely continue aerobic metabolisms at nanomolar O_2 concentrations, allowing an incredibly wide habitat range (e.g., sediments, biofilms, marshes). *Nitrotoga* may also use alternative metabolisms (e.g., sulfur oxidation; see below) in low oxygen environments similar to *Nitrococcus mobilis* (Füssel *et al.*, 2017). The *Nitrotoga* genomes had several O_2 binding proteins including protoglobin, hemerythrin, and potentially nitric oxide dioxygenase, which may facilitate survival under low oxygen conditions.

Intriguingly, the SPKER genome contained an additional *bd*-type terminal cytochrome *c* oxidase, which was also observed in the genomes of two close relatives: *Sideroxydans lithotrophicus* ES-1 and *Gallionella capsiferriformans* ES-2 (Emerson *et al.*, 2013). This terminal oxidase may serve an additional role in energy conservation from organic carbon sources (see below), as *bd*-type terminal oxidases only receive electrons from the quinone pool, while *cbb*$_3$-type terminal oxidases can receive electrons from cytochromes or the quinone pool (Borisov *et al.*, 2011; Morris and Schmidt, 2013).

An F-type ATPase (Complex V) was present in all four *Nitrotoga* genomes for ATP generation via the proton motive force. The periplasmic oxidation of nitrite contributes to the proton gradient, as two protons are released in the reaction. Inorganic phosphate for ATP generation can be stored as polyphosphate in *Nitrotoga* and released by polyphosphate kinase and inorganic pyrophosphatase enzymes.

When nitrite is the only energy source, *Nitrotoga* must generate NADH via reverse electron flow to support normal cellular processes. A canonical cytochrome *bc*$_1$ (Complex III), typically found in proteobacteria, was missing from *Nitrotoga* genomes. However, a suite of genes encoding an alternative complex III (*actAB1B2CDEF*) (Refojo *et al.*, 2012) was found in all *Nitrotoga* genomes and in the related iron-oxidizing bacteria of the Gallionellaceae family (Emerson *et al.*, 2013). Electrons are distributed from the alternative complex III via quinones to succinate dehydrogenase (Complex II) for biochemical intermediate production, or to NADH dehydrogenase (Complex I) for the reduction of NAD^+ (Supplemental Figure 3).

In addition to nitrite oxidation, NXR mediated nitrate reduction with electrons derived from alternative donors has been observed in members of the *Nitrobacter* and *Nitrospira* when oxygen is excluded (Bock *et al*., 1990; Koch *et al*., 2015; Sundermeyer-Klinger *et al*., 1984). Given the genomic potential for survival in low-oxygen environments, we predict that *Nitrotoga* are also capable of nitrate reduction via NXR when oxygen is not available. Alternative electron donors including organic carbon, reduced sulfur compounds, or hydrogen gas (see below) could contribute to the electron transport with NXR functionally replacing the terminal oxidase (Supplemental Figure 3).

Sulfur energetics

Genomic pathways indicated that *Nitrotoga* may be capable of sulfur oxidation. The recently described *Nitrococcus mobilis* Nb-231 genome encoded genes for aerobic sulfur oxidation, and their activity was confirmed in pure culture (Füssel *et al.*, 2017). *Nitrotoga* could likely harness the same sulfur sources using a periplasmic sulfite dehydrogenase (SOR) to oxidize sulfite (SO_3^-) to sulfate (SO_4^{2-}) with electron donation to cytochrome *c*, and a sulfide:quinone oxidoreductase (SQR) that couples hydrogen sulfide (H_2S) oxidation to elemental sulfur (S^0) and reduction of quinone (Supplemental Figure 3).

Hydrogen energetics

The catalytic subunit of a Group 3d [NiFe]-hydrogenase was identified in the CP45, LAW, and MKT genomes with the HydDB online tool (Søndergaard *et al.*, 2016). All other requisite genes were found nearby in the respective genomes. Group 3d [NiFe]-hydrogenases typically act as an NADH oxidoreductase, reducing NAD^+ to NADH with concomitant oxidation of H_2 to water (Peters *et al.*, 2015). Many sequenced NOB have a Group 3b [NiFe]-hydrogenase (Daims *et al.*, 2015; Füssel *et al.*, 2017; Lücker *et al.*, 2013) that is also capable of elemental sulfur or polysulfide reduction to H_2S, an advantageous enzyme that is lacking in *Nitrotoga* genomes. *Nitrospira moscoviensis* utilizes a Group 2a [NiFe]-hydrogenase, and growth on H_2 was confirmed in culture (Koch *et al.*, 2016). Hydrogen oxidation could serve as yet another energy metabolism for *Nitrotoga* cells.

Nitrotoga carbon metabolism

Carbon fixation

Nitrotoga genomes had genes for the complete Calvin cycle (Supplemental Note), supporting CO_2 fixation for autotrophic growth. A sedoheptulose-bisphosphatase gene, an important intermediate enzyme needed to regenerate ribulose-1,5-bisphosphate, was missing but is also missing in the NOB *Nitrobacter winogradskyi*, and the ammonia-oxidizing betaproteobacteria *Nitrosomonas europaea*, and *Nitrosospira multiformis* (Chain *et al.*, 2003; Norton *et al.*, 2008; Starkenburg *et al.*, 2006). Fructose 1,6-bisphosphatase, which typically plays a role in gluconeogenesis, is thought to fill the same role in organisms lacking sedoheptulose-bisphosphatase (Wei *et al.*, 2004; Yoo and Bowien, 1995) and was present in *Nitrotoga* genomes. *Nitrolancea*, *Nitrococcus*, and *Nitrobacter* also utilize the Calvin cycle (Sorokin *et al.*, 2012; Füssel *et al.*, 2017; Starkenburg *et al.*, 2008).

Organic carbon utilization

Nitrotoga genomes encoded all genes for glycolysis, gluconeogenesis, the pentose phosphate pathway, and the tricarboxylic acid (TCA) cycle (Figure 5a). All necessary biochemical intermediates could be generated via these pathways, whether carbon is fixed via the Calvin cycle or imported into the cell. *Nitrotoga* genomes had genes for polysaccharide storage and exopolysaccharide synthesis, which may facilitate growth in biofilms (Supplemental Note).

Various sugar transporters are present in many NOB, including *Nitrococcus mobilis*, *Nitrospira inopinata*, *Nitrospina gracilis*, and *Nitrolancea hollandica* (Füssel *et al.*, 2017; Daims *et al.*, 2015; Lücker *et al.*, 2013; Sorokin *et al.*, 2012), but none were identified in *Nitrotoga* genomes. However, several genes involved in the phosphotransferase system (PTS) to import and phosphorylate sugars were found. The PTS was also found in *Nitrobacter winogradskyi* and *Thiocapsa* sp. KS1 (Starkenburg *et al.*, 2006; Hemp *et al.*, 2016). EI and HPr subunits of the PTS were both found in *Nitrotoga*, but a complete set of substrate-specific EIIA, EIIB, and EIIC components were not identified. Only ascorbate-specific EIIA and nitrogen regulatory EIIA subunits were found. *Nitrotoga* genomes did possess relatives of a recently purified ABC transporter (AfuABC) that was found to import phosphorylated sugars instead of iron as originally reported (Sit *et al.*, 2015).

Nitrotoga genomes had genes for an acetyl-CoA synthetase, which produces acetyl-CoA from acetate, suggesting the use of small organic carbon molecules if they can enter the cytosol. The incomplete *Nitrotoga* genomes, as well as otherwise complete organic carbon oxidation pathways, leaves potential for future identification of organic carbon transport. Preliminary

448 physiology tests showed a dramatically increased rate of nitrite oxidation when acetate and
449 dextrose were provided to the cultures (data not shown).
450
451 ***Nitrotoga* iron acquisition**
452 Iron is often a limiting nutrient in oligotrophic freshwater environments. *Nitrotoga*
453 genomes encoded an impressive array of genes useful for iron scavenging, including a complete
454 TonB-dependent transport system with up to four copies of the outer membrane transport
455 energization protein complex (ExbB/ExbD/TonB) and up to 16 TonB-dependent outer
456 membrane siderophore receptors (CP45: 10; LAW: 7; MKT: 6; SPKER: 16). All of the TonB-
457 dependent transporters (TBDTs) fell within the CirA and/or Fiu superfamilies (based on
458 BLASTP and conserved domain results), which have been shown to transport a wide variety of
459 siderophores including citrate, aerobactin, enterobactin, and salmochelin (Porcheron et al., 2013).
460 Eight TBDTs (MKT: 0; CP45: 1; LAW: 1; SPKER: 5) were preceded by FecI-FecR transcription
461 factors known to monitor ferric dicitrate concentrations within the cell and control transcription
462 of TBDTs in *E. coli* (Braun *et al.*, 2003). Siderophore synthesis seems unlikely within *Nitrotoga*
463 (based on gene predictions; Supplemental Note), however *Nitrotoga* may scavenge siderophores
464 from other bacteria within the community. Additionally, several iron pumps and iron storage
465 proteins were present (Supplemental Note). Iron is a critical cofactor in the NXR enzyme and the
466 varied iron acquisition methods of *Nitrotoga* may offer a competitive advantage for these
467 organisms in iron-limited environments.
468
469 ***Nitrotoga* heavy metal transport and defense**
470 Molybdate (MoO_4^{2-}) transport genes (*modABC*) were found in the CP45, LAW, and
471 MKT genomes; while the SPKER genome had two *modA* copies but no *modB* or *modC*.
472 However, the SPKER genome encoded a pair of molybdenum storage genes (*mosAB*) for
473 intracellular storage of molybdenum (Fenske *et al.*, 2005), not found in other *Nitrotoga*.
474 Molybdate is the naturally occurring form of molybdenum, which is a necessary cofactor for
475 NXR, so *Nitrotoga* cells must have at least one method of acquiring molybdenum.
476 Heavy metal efflux systems were prevalent in all *Nitrotoga* genomes. An *apaG* cobalt
477 and magnesium efflux protein, as well as complete cobalt-zinc-cadmium resistance transporters
478 (*czcABC*) were found in all genomes. A chromate transporter and divalent cation tolerance
479 protein may help protect against heavy metal accumulation in *Nitrotoga*.
480 Arsenate (AsO_4^{3-}) can enter cells through normal phosphate transport systems, and
481 *Nitrotoga* had two different mechanisms for arsenate removal (Figure 5b). The LAW and
482 SPKER genomes had *arsC* and *arsB* genes responsible for reducing arsenate to arsenite (AsO_3^{3-}),
483 and pumping arsenite from the cytoplasm, respectively. The CP45 and MKT genomes encoded
484 an additional subunit, *arsA*, which acts as an ATPase to provide energy for export, while the
485 ArsB pump can work alone with energy from the proton motive force.
486 All *Nitrotoga* genomes encoded the alpha (*cusA*) and beta (*cusB*) portions of the Cus
487 Cu(I)/Ag(I) efflux system, composing the inner membrane transporter and periplasmic
488 membrane fusion protein, respectively. The outer membrane factor channel protein, *cusC*, was
489 only found in a complete operon in the LAW genome along with the *cusF* metallochaperone and
490 *cusR-cusS* two-component signaling system. The MKT and SPKER genomes had *cusR* and *cusS*
491 genes preceding a TolC-like outer membrane protein that could be CusC, but these were found
492 separate from *cusAB*. A second copper exporter belonging to the Cue system was found in the
493 LAW, MKT, and SPKER genomes consisting of a Cu^{2+} exporter (*copA*) and multicopper oxidase

(*cueO*). The CP45 genome alone had a *copZ* gene that serves as a chaperone for Cu^{+} export via CopA.

Chromate (CrO_4^{2-}) can enter cells via normal sulfate uptake systems. A putative soluble chromate reductase was found in all *Nitrotoga* genomes that can reduce the Cr(VI) found in chromate to the less toxic Cr(III), producing harmful reactive oxygen species (ROS) in the process which must be dealt with by other defenses (Thatoi *et al*., 2014). In addition, a chromate transporter gene (*chrA*) was found in all genomes for the excretion of chromate ions from the cytoplasm.

A variety of other heavy metal defenses were found in individual *Nitrotoga* genomes (Figure 5b), for instance, the CP45 genome encoded a gold/copper resistance efflux pump belonging to the resistance nodulation division (RND) superfamily. The LAW genome encoded a mercury resistance system (*merRTPFA*) for Hg(II) transport and reduction to volatile Hg(0), and the LAW and SPKER genomes encoded *terC*-like proteins associated with tellurite resistance but appear to be lacking other required genes for tellurite efflux and/or reduction. Heavy metal transport is critical for cofactor acquisition (e.g., iron, molybdenum) and for defense against metals that are toxic at even low concentrations (e.g., mercury, arsenic). *Nitrotoga* genomes possessed varied metal transporters, which likely help support cells in the contaminated rivers sampled in this study.

Other defense mechanisms in *Nitrotoga*

In support of the aerobic metabolisms of *Nitrotoga*, all four genomes contained necessary catalase, superoxide dismutase (Cu-Zn and Fe-Mn families), and peroxiredoxin genes to combat reactive oxygen species (ROS) (Figure 5). Cytochrome *c* peroxidase genes were found in the CP45, LAW, and MKT genomes to intercept peroxides in the periplasm with electrons donated from cytochrome *c*. Additionally, a rubrerythrin protein used to combat ROS specifically in anaerobic bacteria was found, a reminder of the likely microaerophilic or anaerobic ancestry of *Nitrotoga* within the Gallionellaceae.

A few antibiotic resistance genes were present among the *Nitrotoga* genomes, including a generic antibiotic biosynthesis monooxygenase, VanZ-like family proteins that may confer low-level antibiotic resistance, and a broad-spectrum multidrug efflux system (AcrAB-TolC). An erythromycin esterase homolog was found in the LAW genome, which may confer resistance to erythromycin. Preliminary physiology tests indicated that *Nitrotoga* continued nitrite oxidation in the presence of various antibiotics (data not shown).

The SPKER *Nitrotoga* genome possessed an interesting suite of genes responsible for DNA phosphorothionation (*iscS*, *dndBCDE*) (swapping a non-bridging oxygen atom for a sulfur atom) that offers resistance to a related restriction-modification system (*dptHGF*) (Xu *et al*., 2010) found nearby on the genome. DNA phosphorothionation was recently found to expand microbial growth range under multiple stresses due to its intrinsic antioxidant function (Yang *et al.*, 2017). This ability may greatly expand the environments in which *Ca.* Nitrotoga coloradensis SPKER can proliferate, adding to its already wide potential range.

Motility and chemotaxis

All *Nitrotoga* genomes carried genes necessary for flagellar assembly and operation, as well as signal transducing pathways to stimulate gliding or twitching motility via a type IV pilus assembly. All genomes had general chemotaxis genes, but the SPKER genome was notably missing an aerotaxis receptor (*aer*) capable of detecting oxygen concentrations. If *Nitrotoga* are

indeed motile, they may migrate towards areas of greater reducing potential, and likely lose external motile structures when growing in biofilms or in culture, as the presence of flagella or pili have not been noted in previous studies (Alawi *et al.*, 2007, 2009; Hüpeden *et al.*, 2016; Lücker *et al.*, 2015; Ishii *et al.*, 2017).

The presence of a single acyl homoserine lactone synthase (LuxI homolog) in each genome, as well as several LuxR family transcriptional receptors and multiple homoserine efflux transporters (RhtA), indicate that *Nitrotoga* may use quorum sensing to communicate. Other NOB have been documented to use quorum sensing at high cell densities and the *Nitrotoga* LuxI homolog grouped phylogenetically with other sequenced Betaproteobacteria (Sayavedra-Soto *et al.*, 2015; Mellbye *et al.*, 2017). *Nitrotoga* have been documented to grow in flocs and were shown to co-aggregate with AOB in WWTPs (Alawi *et al*., 2007; Lücker *et al*., 2015), so they may use quorum sensing to form microcolonies in response to their environment.

Environmental distribution

Nitrotoga-like 16S rRNA sequences were present in 2,410 different samples (from 183,154 total SRA runs analyzed with IMNGS) (Figure 6). *Nitrotoga* containing samples comprised 70 different user-defined environments with a global distribution across seven continents spanning from the tropics to the poles (Figure 6C). About 9% of all freshwater environments (517/5,678) had *Nitrotoga*-like OTUs, with some communities consisting of nearly 10% relative abundance of *Nitrotoga*-like OTUs. Nearly half of all wetland samples (24/49) had *Nitrotoga*-like OTUs with an average relative abundance of 0.89%, however most of these samples were part of the same BioProject. A high proportion of activated sludge (15%) and wastewater (17%) samples also harbored relevant populations of *Nitrotoga*-like OTUs (0.55% average relative abundance), similar to previous reports (Lücker et al., 2015). The relative abundance of *Nitrotoga*-like OTUs ranged from 0.02%-13.03% across 540 soil samples. Of the few (6/6,233) marine samples that had *Nitrotoga*-like OTUs, most were found near freshwater rivers or wetlands.

A few samples in the SRA dataset had very high relative abundance of *Nitrotoga*-like OTUs compared to other samples. Several of these samples were associated with enrichment cultures from their respective environments (e.g., biofilm, hydrocarbon, and other). For example, a series of samples from a single study which involved enrichment from waterworks sand filters in the Netherlands had *Nitrotoga*-like OTUs comprising as much as 61% of the microbial community.

Nitrotoga distribution was also evaluated across 80 water column and sediment samples from three rivers in Colorado (Bear Creek, Cherry Creek, and South Platte River), based on amplification with general 16S rRNA gene primers targeting the total bacterial community (Supplemental Figure S4). *Nitrotoga*-like OTUs were identified in 85% of the samples, and *Nitrospira*-like OTUs were identified in 98% of the samples. Both groups co-occurred in 85% of the samples. No other NOB sequences were identified. Within each individual sample, the summed relative abundance ranged from 0-4.5% of the total bacterial community for *Nitrotoga*-like OTUs and 0-6.4% for *Nitrospira*-like OTUs. The summed relative abundance of *Nitrotoga*-like OTUs was greater than that of *Nitrospira*-like OTUs in 21% of the samples. *Nitrobacter* and *Nitrospira* are typically the more commonly studied NOB in freshwater systems (Daims *et al.*, 2016; Cai *et al.*, 2018), but future efforts should now also consider *Nitrotoga* as potential important players in freshwater nitrite oxidation.

The *Nitrotoga* species presented here were enriched from the water column and sediments of urban- and agriculturally-impacted rivers, suggesting a broad habitat range within freshwater systems and a greater role in freshwater nitrification than previously expected. Sampling for these *Nitrotoga* enrichment cultures occurred in winter and spring when *in situ* water temperatures are typically <15°C and cultures were grown with 300 µM nitrite. All other reported *Nitrotoga* were enriched at low temperatures (4-17°C) and low nitrite concentration (300 µM) (Alawi *et al.*, 2007, 2009; Hüpeden *et al.*, 2016; Ishii *et al.*, 2017). Temperature and substrate availability may play a role in niche differentiation of NOB, with *Nitrotoga* increasing in abundance at colder temperatures and lower nitrite concentrations, but further molecular and physiology experiments would be needed to confirm this.

CONCLUSIONS

The enrichment of four novel *Nitrotoga* species, characterization of the first *Nitrotoga* genomes, and analysis of the distribution of *Nitrotoga* 16S rRNA sequences has expanded our knowledge of NOB ecology. The divergent NXR enzyme may indicate a novel evolution of nitrite oxidation in *Nitrotoga* separate from other known NOB. *Nitrotoga* have an exceptionally diverse metabolic profile, likely allowing proliferation under variable nutrient and oxygen conditions. The prevalence of *Nitrotoga*-like sequences found in globally distributed habitats (e.g., freshwater, soil, wetland, wastewater) suggests that *Nitrotoga* likely play a previously underappreciated functional role on a global scale. Future work should determine *Nitrotoga* ecophysiology, including nitrite oxidation kinetics and optimal growth conditions. Determining the functional redundancy and niche differentiation of *Nitrotoga* compared to other NOB will be important for understanding their role in nitrite oxidation under varying environmental conditions. Understanding the sensitivity or resilience of *Nitrotoga* to disturbances will be important for predicting their response to environmental change.

ACKNOWLEGEMENTS

We thank Adrienne Narrowe and Christopher Miller for guidance on bioinformatic data analyses. We would also like to thank Hannah Clark, Nicklaus Deevers, Colin Beacom, Michael Kain, and Munira Lantz for their assistance in cultivation and preliminary physiology experiments and Sladjana Subotic and Anna Scopp for their help with environmental distribution sample collection and processing.

COMPETING INTERESTS

The authors declare no competing interests. Portions of this manuscript were previously published as a part of University of Colorado Denver Master's thesis submission (AB, 2017).
Funding was provided by the University of Colorado, Denver and the City and County of Denver.

SUPPLEMENTARY INFORMATION

Supplementary Information accompanies this paper.

626 **REFERENCES**
627
628 Alawi M, Lipski A, Sanders T, Pfeiffer EM, Spieck E. (2007). Cultivation of a novel cold-
629 adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J* **1**: 256–264.
630 Alawi M, Off S, Kaya M, Spieck E. (2009). Temperature influences the population structure of
631 nitrite-oxidizing bacteria in activated sludge. *Environ Microbiol Rep* **1**: 184–190.
632 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* (2012). SPAdes:
633 A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput*
634 *Biol* **19**: 455–477.
635 Bartosch S, Wolgast I, Spieck E, Bock E. (1999). Identification of Nitrite-Oxidizing Bacteria
636 with Monoclonal Antibodies Recognizing the Nitrite Oxidoreductase. *Appl Environ Microbiol*
637 **65**: 4126–4133.
638 Bereschenko LA, Stams AJM, Euverink GJW, Van Loosdrecht MCM. (2010). Biofilm formation
639 on reverse osmosis membranes is initiated and dominated by Sphingomonas spp. *Appl Environ*
640 *Microbiol* **76**: 2623–2632.
641 Bock E, Koops H-P, Möller UC, Rudert M. (1990). A new facultatively nitrite oxidizing
642 bacterium, Nitrobacter vulgaris sp. nov. *Arch Microbiol* **153**: 105–110.
643 Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. (2011). The cyotchrome bd respiratory
644 oxygen reductases. *Biochim Biophys Acta - Bioenerg* 1398–1413.
645 Braun V, Mahren S, Ogierman M. (2003). Regulation of the Fecl-type ECF sigma factor by
646 transmembrane signalling. *Curr Opin Microbiol* **6**: 173–180.
647 Cai M, Ng S-K, Lim CK, Lu H, Jia Y, Lee PKH. (2018). Physiological and Metagenomic
648 Characterizations of the Synergistic Relationships between Ammonia- and Nitrite-Oxidizing
649 Bacteria in Freshwater Nitrification. *Front Microbiol* **9**: 1–13.
650 Caro-Quintero A, Konstantinidis KT. (2012). Bacterial species may exist, metagenomics reveal.
651 *Environ Microbiol* **14**: 347–355.
652 Chain P, Lamerdin J, Larimer F, Regala W, Lao V, Land M, *et al.* (2003). Complete genome
653 sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas
654 europaea. *J Bacteriol* **185**: 2759–73.
655 Chen Y, Wu L, Boden R, Hillebrand A, Kumaresan D, Moussard H, *et al.* (2009). Life without
656 light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in
657 Movile Cave. *ISME J* **3**: 1093–1104.
658 Christner BC, Priscu JC, Achberger AM, Barbante C, Carter SP, Christianson K, *et al.* (2014). A
659 microbial ecosystem beneath the West Antarctic ice sheet. *Nature* **512**: 310–313.
660 Daims H, Lebedeva E V., Pjevac P, Han P, Herbold C, Albertsen M, *et al.* (2015). Complete
661 nitrification by Nitrospira bacteria. *Nature*. e-pub ahead of print, doi: 10.1038/nature16461.
662 Daims H, Lücker S, Wagner M. (2016). A New Perspective on Microbes Formerly Known as
663 Nitrite-Oxidizing Bacteria. *Trends Microbiol* **24**: 699–712.
664 Darling AE, Mau B, Perna NT. (2010). progressiveMauve : Multiple Genome Alignment with
665 Gene Gain , Loss and Rearrangement. **5**. e-pub ahead of print, doi:
666 10.1371/journal.pone.0011147.
667 Emerson D, Field EK, Chertkov O, Davenport KW, Goodwin L, Munk C, *et al.* (2013).
668 Comparative genomics of freshwater Fe-oxidizing bacteria: Implications for physiology,
669 ecology, and systematics. *Front Microbiol* **4**: 1–17.
670 Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, *et al.* (2015). Anvi'o: an
671 advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: e1319.

672 Fan L, Song C, Meng S, Qiu L, Zheng Y, Wu W, *et al.* (2016). Spatial distribution of planktonic
673 bacterial and archaeal communities in the upper section of the tidal reach in Yangtze River. *Sci*
674 *Rep* **6**: 39147.
675 Fenske D, Gnida M, Schneider K, Meyer-Klaucke W, Schemberg J, Henschel V, *et al.* (2005). A
676 new type of metalloprotein: The Mo storage protein from Azotobacter vinelandii contains a
677 polynuclear molybdenum-oxide cluster. *ChemBioChem* **6**: 405–413.
678 Figdore BA, Stensel HD, Winkler M-KH. (2017). Comparison of different aerobic granular
679 sludge types for activated sludge nitrification bioaugmentation potential. *Bioresour Technol* **251**:
680 189–196.
681 Füssel J, Lam P, Lavik G, Jensen MM, Holtappels M, Günter M, *et al.* (2012). Nitrite oxidation
682 in the Namibian oxygen minimum zone. *ISME J* **6**: 1200–1209.
683 Füssel J, Lücker S, Yilmaz P, Nowka B, van Kessel MAHJ, Bourceau P, *et al.* (2017).
684 Adaptability as the key to success for the ubiquitous marine nitrite oxidizer *Nitrococcus*. *Sci Adv*
685 **3**: e1700807.
686 Gardner PR, Gardner AM, Martin LA, Salzman AL. (1998). Nitric oxide dioxygenase: an
687 enzymic function for flavohemoglobin. *Proc Natl Acad Sci U S A* **95**: 10378–10383.
688 Griess-Romijn van Eck E. (1966). Physiological and chemical tests for drinking water.
689 (Nederlands Normalisatie Instituut).
690 Han H, Hemp J, Pace LA, Ouyang H, Ganesan K, Roh JH, *et al.* (2011). Adaptation of aerobic
691 respiration to low O2 environments. *Proc Natl Acad Sci* **108**: 14109–14114.
692 Hemp J, Lücker S, Schott J, Pace LA, Johnson JE, Schink B, *et al.* (2016). Genomics of a
693 phototrophic nitrite oxidizer: insights into the evolution of photosynthesis and nitrification. *ISME*
694 *J* **10**: 1–10.
695 Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, *et al.*
696 (2015). The Standard Operating Procedure of the DOE-JGI Microbial Genome Annotation
697 Pipeline (MGAP v 4). *Stand Genomic Sci* **10**: 1–6.
698 Hüpeden J, Wegen S, Off S, Lücker S, Bedarf Y, Daims H, *et al.* (2016). Relative Abundance of
699 Nitrotoga spp. in a biofilter of a cold freshwater aquaculture plant appears to be stimulated by a
700 slightly acidic pH-value. *Appl Environ Microbiol* **82**: 1838–1845.
701 Ishii K, Fujitani H, Soh K, Nakagawa T, Takahashi R, Tsuneda S. (2017). Enrichment and
702 Physiological Characterization of a Cold-Adapted Nitrite Oxidizer Nitrotoga sp. from Eelgrass
703 Sediments. *Appl Environ Microbiol* **83**: 1–14.
704 Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. (2017). High-throughput
705 ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *bioRxiv* 225342.
706 Kang DD, Froula J, Egan R, Wang Z. (2015). MetaBAT, an efficient tool for accurately
707 reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.
708 Karkman A, Mattila K, Tamminen M, Virta M. (2011). Cold temperature decreases bacterial
709 species richness in nitrogen-removing bioreactors treating inorganic mine waters. *Biotechnol*
710 *Bioeng* **108**: 2876–2883.
711 Katoh K, Standley DM. (2013). MAFFT multiple sequence alignment software version 7:
712 Improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.
713 van Kessel MAHJ, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, *et al.*
714 (2015). Complete nitrification by a single microorganism. *Nature* **528**: 555–559.
715 Kinnunen M, Gulay A, Albrechtsen H-J, Dechesne A, Smets BF. (2017). Nitrotoga is selected
716 over Nitrospira in newly assembled biofilm communities from a tap water source community at
717 increased nitrite loading. *Environ Microbiol* **0**: 1–41.

718 Kits KD, Sedlacek CJ, Lebedeva E V., Han P, Bulaev A, Pjevac P, *et al.* (2017). Kinetic analysis
719 of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**: 269–272.
720 Kloer DP, Hagel C, Heider J, Schulz GE. (2006). Crystal Structure of Ethylbenzene
721 Dehydrogenase from Aromatoleum aromaticum. *Structure* **14**: 1377–1388.
722 Koch H, Galushko A, Albertsen M, Schintlmeister A, Spieck E, Richter A, *et al.* (2016). Growth
723 of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Reasearch | Reports* **1696**: 761–763.
724 Koch H, Lücker S, Albertsen M, Kitzinger K, Herbold C, Spieck E, *et al.* (2015). Expanded
725 metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc Natl*
726 *Acad Sci* **112**: 11371–11376.
727 Konstantinidis KT, Rosselló-móra R, Amann R. (2017). Uncultivated microbes in need of their
728 own taxonomy. *ISME J* **11**: 2399–2406.
729 Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, *et al.* (2016). IMNGS:
730 A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and
731 diversity studies. *Sci Rep* **6**: 1–9.
732 Lebedeva E V., Alawi M, Jozsa PG, Daims H, Spieck E. (2008). Physiological and phylogenetic
733 characterization of a novel lithoautotrophic nitrite-oxidizing bacterium, 'Candidatus Nitrospira
734 bockiana'. *Int J Syst Evol Microbiol* **58**: 242–250.
735 Letunic I, Bork P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and
736 annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.
737 Li D, Liu CM, Luo R, Sadakane K, Lam TW. (2014). MEGAHIT: An ultra-fast single-node
738 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
739 *Bioinformatics* **31**: 1674–1676.
740 Li D, Qi R, Yang M, Zhang Y, Yu T. (2011). Bacterial community characteristics under long-
741 term antibiotic selection pressures. *Water Res* **45**: 6063–6073.
742 Lücker S, Nowka B, Rattei T, Spieck E, Daims H. (2013). The Genome of Nitrospina gracilis
743 Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front Microbiol*
744 **4**: 1–19.
745 Lücker S, Schwarz J, Gruber-Dorninger C, Spieck E, Wagner M, Daims H. (2015). Nitrotoga-
746 like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment
747 plants. *ISME J* **9**: 708–720.
748 Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, *et al.* (2010). A Nitrospira
749 metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing
750 bacteria. *Proc Natl Acad Sci U S A* **107**: 13479–13484.
751 Martinez-espinosa RM, Dridge EJ, Bonete MJ, Butt JN, Butler CS, Sargent F, *et al.* (2007).
752 Look on the positive side ! The orientation , identification and bioenergetics of 'Archaeal'
753 membrane-bound nitrate reductases. *FEMS Microbiol Lett* **276**: 129–139.
754 Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial
755 Beta-diversity depend on spatial scale. *Proc Natl Acad Sci* **108**: 7850–7854.
756 Mellbye BL, Spieck E, Bottomley PJ, Sayavedra-Soto LA. (2017). Acyl-Homoserine Lactone
757 Production in Nitrifying Bacteria of the Genera Nitrosopira, Nitrobacter, and Nitrospria
758 Identified via a Survey of Putative Quorum-Sensing Genes. *Appl Environ Microbiol* **83**: 1–13.
759 Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF, Pace N, *et al.* (2011). EMIRGE:
760 reconstruction of full-length ribosomal genes from microbial community short read sequencing
761 data. *Genome Biol* **12**: R44.
762 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. (2007). KAAS: An automatic genome
763 annotation and pathway reconstruction server. *Nucleic Acids Res* **35**: 182–185.

764 Morris RL, Schmidt TM. (2013). Shallow breathing: bacterial life at low O2. *Nat Rev Microbiol*
765 **11**: 205–212.
766 Na H, Kim OS, Yoon SH, Kim Y, Chun J. (2011). Comparative approach to capture bacterial
767 diversity of coastal waters. *J Microbiol* **49**: 729–740.
768 Ngugi DK, Blom J, Stepanauskas R, Stingl U. (2015). Diversification and niche adaptations of
769 Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *ISME J* **10**: 1383–1399.
770 Norton JM, Klotz MG, Stein LY, Arp DJ, Bottomley PJ, Chain PSG, *et al.* (2008). Complete
771 Genome Sequence of Nitrosospira multiformis, an Ammonia-Oxidizing Bacterium from the Soil
772 Environment. *Appl Environ Microbiol* **74**: 3559–3572.
773 Nowka B, Daims H, Spieck E. (2015). Comparison of Oxidation Kinetics of Nitrite-Oxidizing
774 Bacteria: Nitrite Availability as a Key Factor in Niche Differentiation. *Appl Environ Microbiol*
775 **81**: 745–753.
776 Pachiadaki MG, Sintes E, Bergauer K, Brown JM, Record NR, Swan BK, *et al.* (2017). Major
777 role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science (80-)* **358**: 1046–1051.
778 Parada AE, Needham DM, Fuhrman JA. (2015). Every base matters: Assessing small subunit
779 rRNA primers for marine microbiomes with mock communities, time series and global field
780 samples. *Environ Microbiol*. e-pub ahead of print, doi: 10.1111/1462-2920.13023.
781 Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing
782 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
783 *Genome Res* **25**: 1043–55.
784 Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, *et al.* (2015). [FeFe]-
785 and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochim Biophys Acta - Mol Cell*
786 *Res* **1853**: 1350–1369.
787 Pinto AJ, Marcus DN, Ijaz Z, Bautista-de los Santos QM, Dick GJ, Raskin L. (2015).
788 Metagenomic Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking
789 Water System. *mSphere* **1**: e00054-15.
790 Porcheron G, Garénaux A, Proulx J, Sabri M, Dozois CM. (2013). Iron, copper, zinc, and
791 manganese transport and regulation in pathogenic Enterobacteria: correlations between strains,
792 site of infection and the relative importance of the different metal transport systems for
793 virulence. *Front Cell Infect Microbiol* **3**: 1–24.
794 Pradhan S, Srinivas TNR, Pindi PK, Kishore HH, Begum Z, Singh PK, *et al.* (2010). Bacterial
795 biodiversity from Roopkund Glacier, Himalayan mountain ranges, India. *Extremophiles* **14**: 377–
796 395.
797 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2013). The SILVA
798 ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic*
799 *Acids Res* **41**: 590–596.
800 Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. (2011). Removing Noise From
801 Pyrosequenced Amplicons. *BMC Bioinformatics* **12**. e-pub ahead of print, doi:
802 10.1128/JVI.02271-09.
803 Ramanathan B, Boddicker AM, Roane TM, Mosier AC. (2017). Nitrifier Gene Abundance and
804 Diversity in Sediments Impacted by Acid Mine Drainage. *Front Microbiol* **8**: 1–16.
805 Refojo PN, Teixeira M, Pereira MM. (2012). The Alternative complex III: Properties and
806 possible mechanisms for electron transfer and energy conservation. *BBA - Bioenerg* **1817**: 1852–
807 1859.
808 Roden EE, McBeth JM, Blöthe M, Percak-Dennett EM, Fleming EJ, Holyoke RR, *et al.* (2012).
809 The microbial ferrous wheel in a neutral pH groundwater seep. *Front Microbiol* **3**: 1–18.

Rodriguez-R LM, Konstantinidis KT. (2014). Bypassing Cultivation To Identify Bacterial Species. *Microbe Mag* **9**: 111–118.

Rodriguez-R LM, Konstantinidis KT. (2016). The enveomics collection : a toolbox for specialized analyses of microbial genomes and metagenomes. *Peer J Prepr*. e-pub ahead of print, doi: 10.7287/peerj.preprints.1900v1.

Sargent F. (2007). The twin-arginine transport system: moving folded proteins across membranes. *Biochem Soc Trans* **35**: 835–847.

Sattin SR, Cleveland CC, Hood E, Reed SC, King AJ, Schmidt SK, *et al.* (2009). Functional shifts in unvegetated, perhumid, recently-deglaciated soils do not correlate with shifts in soil bacterial community composition. *J Microbiol* **47**: 673–681.

Saunders AM, Albertsen M, Vollesen J, Nielsen PH. (2015). The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* **10**: 11–20.

Sayavedra-Soto L, Ferrell R, Dobie M, Mellbye B, Chaplen F, Buchanan A, *et al.* (2015). Nitrobacter winogradskyi transcriptomic response to low and high ammonium concentrations. *FEMS Microbiol Lett* **362**: 1–7.

Schmidt SK, Nemergut DR, Miller AE, Freeman KR, King AJ, Seimon A. (2009). Microbial activity and diversity during extreme freeze-thaw cycles in periglacial soils, 5400 m elevation, Cordillera Vilcanota, Peru. *Extremophiles* **13**: 807–816.

Sit B, Crowley SM, Bhullar K, Lai CCL, Tang C, Hooda Y, *et al.* (2015). Active Transport of Phosphorylated Carbohydrates Promotes Intestinal Colonization and Transmission of a Bacterial Pathogen. *PLoS Pathog* **11**: 1–22.

Søndergaard D, Pedersen CNS, Greening C. (2016). HydDB: A web tool for hydrogenase classification and analysis. *Sci Rep* **6**: 1–8.

Sorokin DY, Lücker S, Vejmelkova D, Kostrikina N a, Kleerebezem R, Rijpstra WIC, *et al.* (2012). Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J* **6**: 2245–2256.

Spieck E, Aamand J, Bartosch S, Bock E. (1996). Immunocytochemical detection and location of the membrane-bound nitrite oxidoreductase in cells of Nitrobacter and Nitrospira. *FEMS Microbiol Lett* **139**: 71–76.

Spieck E, Ehrich S, Aamand J, Bock E. (1998). Isolation and immunocytochemical location of the nitrite-oxidizing system in Nitrospira moscoviensis. *Arch Microbiol* **169**: 225–230.

Srinivas TNR, Singh SM, Pradhan S, Pratibha MS, Kishore KH, Singh AK, *et al.* (2011). Comparison of bacterial diversity in proglacial soil from Kafni Glacier, Himalayan Mountain ranges, India, with the bacterial diversity of other glaciers in the world. *Extremophiles* **15**: 673–690.

Starkenburg SR, Chain PSG, Sayavedra-Soto LA, Hauser L, Land ML, Larimer FW, *et al.* (2006). Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium Nitrobacter winogradskyi Nb-255. *Appl Environ Microbiol* **72**: 2050–2063.

Starkenburg SR, Larimer FW, Stein LY, Klotz MG, Chain PSG, Sayavedra-Soto LA, *et al.* (2008). Complete genome sequence of Nitrobacter hamburgensis X14 and comparative genomic analysis of species within the genus Nitrobacter. *Appl Environ Microbiol* **74**: 2852–63.

Sundermeyer-Klinger H, Meyer W, Warninghoff B, Bock E. (1984). Membrane-bound nitrite oxidoreductase of Nitrobacter: evidence for a nitrate reductase system. *Arch Microbiol* **140**: 153–158.

Thatoi H, Das S, Mishra J, Rath BP, Das N. (2014). Bacterial chromate reducatase, a potential enzyme for bioremediation of hexavalent chromium: A review. *J Environ Manage* **146**: 383–

399.
Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U. (1986). Nitrospira marina gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* **144**: 1–7.
Watson SW, Waterbury JB. (1971). Characteristics of two marine nitrite oxidizing bacteria, Nitrospina gracilis nov. gen. nov. sp. and Nitrococcus mobilis nov. gen. nov. sp. *Arch Mikrobiol* **77**: 203–230.
Wei X, Sayavedra-Soto LA, Arp DJ. (2004). The transcription of the cbb operon in Nitrosomonas europaea. *Microbiology* **150**: 1869–1879.
White CP, DeBry RW, Lytle DA. (2012). Microbial survey of a full-scale, biologically active filter for treatment of drinking water. *Appl Environ Microbiol* **78**: 6390–6394.
Xu T, Yao F, Zhou X, Deng Z, You D. (2010). A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella. *Nucleic Acids Res* **38**: 7133–7141.
Yang J, Chen S, Li H. (2016). Dewatering sewage sludge by a combination of hydrogen peroxide, jute fiber wastes and cationic polyacrylamide. *Int Biodeterior Biodegrad* 1–7.
Yang Y, Xu G, Liang J, He Y, Xiong L, Li H, *et al.* (2017). DNA Backbone Sulfur-Modification Expands Microbial Growth Range under Multiple Stresses by its anti-oxidation function. *Sci Rep* **7**: 1–9.
Yoo J-G, Bowien B. (1995). Analysis of the cbbF genes from Alcaligenes eutrophus that encode fructose-1,6-/sedoheptulose-1,7-bisphosphatase. *Curr Microbiol* **31**: 55–61.
Ziegler AS, McIlroy SJ, Larsen P, Albertsen M, Hansen AA, Heinen N, *et al.* (2016). Dynamics of the fouling layer microbial community in a membrane bioreactor. *PLoS One* **11**: 1–14.

879 **FIGURE LEGENDS**
880
881 **Table 1.** Average nucleotide identity (ANI) between enriched Nitrotoga genomes and close
882 relative genomes available on NCBI (unhighlighted cells; top right) and average amino acid
883 identity (AAI) pairwise comparisons (highlighted cells; bottom left).
884
885 **Figure 1.** Nitrite consumption by *Nitrotoga* enrichment cultures over time. Each enrichment
886 culture was inoculated in three replicates, and nitrite concentration was quantified
887 colorimetrically in triplicate at each time point. Error bars show the standard deviation of each
888 time point; error bars that appear to be missing are too small to be visualized. Sterile FNOM was
889 used as a control and plotted with each culture. Logarithmic declines in nitrite concentration
890 were used to calculate the nitrite oxidation rate for each biological replicate. The average nitrite
891 consumption per day is shown with standard deviation among triplicate cultures.
892
893 **Figure 2.** Anvi'o pangenome analysis of four *Nitrotoga* genomes. Coding sequences for all four
894 genomes (10,666) in total grouped into 4,001 protein clusters (PCs), based on a pairwise BLAST
895 of all coding sequences from all *Nitrotoga* genomes and clustering using the MCL algorithm
896 (mcl=10). The core genome was represented by 1,803 PCs found in all four *Nitrotoga* genomes.
897 Each individual genome (e.g., MKT only) contained 293-625 unique PCs not found in any other
898 *Nitrotoga* genome.
899
900 **Figure 3. A)** Maximum likelihood phylogenetic tree of 16S rRNA gene sequences from
901 representative *Ca*. Nitrotoga sequences and close relatives. Sequences were aligned across 1,422
902 positions; phylogenetic trees were generated using RAxML (version 8.2.9) with 1,000 rapid
903 bootstraps and the GTRGAMMA model of nucleotide substitution. Nodes with bootstrap support
904 values ≥50% are shown with a black circle. A midpoint root was used when selecting outgroup
905 sequences. Bolded sequence names have been enriched in culture. Nodes with a star represent
906 organisms presented in this study. **B)** BLASTN comparisons of the near full-length 16S rRNA
907 gene sequences from the eight known enriched *Nitrotoga* sp.
908
909 **Figure 4.** Phylogenetic and structural analysis of the alpha subunit of nitrite oxidoreductase
910 (NxrA). **A)** Phylogeny of 122 members of the Type II DMSO reductase protein family aligned
911 and manually trimmed to 1,651 amino acid positions. References were selected to include both
912 cytoplasmic-facing "Low Affinity" and periplasmic-facing "High Affinity" NxrA (Ngugi *et al.*,
913 2015), as well as other family members: PcrA (perchlorate reductase), EbdA (ethylbenzene
914 dehydrogenase), DdhA (dimethylsulfide dehydrogenase), ClrA (chlorate reductase), SerA
915 (selenite reductase), and NarG (nitrate reductase). Putative enzymes are marked as "put.". The
916 number of sequences in collapsed nodes are shown in parentheses. RAxML rapid bootstrap
917 support values ≥90% are marked with a black circle. **B)** Partial alignment of selected important
918 residues in *Nitrotoga*, periplasmic-facing, and cytoplasmic-facing NxrA (including *Thiocapsa*).
919 Highlights represent Tat signal peptides (Orange), Fe-S cluster binding residues (Yellow),
920 molybdenum coordinating residues (Pink), and nitrite/nitrate binding residues (Light Blue).
921

Figure 5. Schematic representation of *Nitrotoga* genome features. **A)** The core *Nitrotoga* genome including predicted functions shared by all four genomes. The question mark located in the NxrC subunit symbolizes the uncertainty in whether or not the holoenzyme is anchored to the cell membrane. **B)** Predicted function of individual genomic features found within one or more genomes, but not shared among all four. Each genome had 178-301 unique hypothetical proteins, and an additional 171 hypothetical proteins were shared among two or three genomes.

Figure 6. Global distribution of *Nitrotoga*-like sequences. OTUs from 16S rRNA gene amplicon studies deposited as SRA runs were clustered by IMNGS. All runs with OTUs ≥ 97% identity to cultured *Nitrotoga* 16S rRNA gene sequences (from this study), an alignment of at least 200 bp, and at least 100 total reads were kept. **A)** The percent of SRA runs with *Nitrotoga*-like OTUs within respective environments are shown with the number of SRA runs with *Nitrotoga*-like OTUs displayed above each bar. **B)** The relative abundance of *Nitrotoga*-like OTUs was averaged across all four queried *Nitrotoga* 16S rRNA gene sequences and plotted by environment. **C)** Global distribution of SRA runs with *Nitrotoga*-like OTUs. 484 of the 2,410 SRA runs with *Nitrotoga*-like OTUs did not have geographic information available, including all of the "Hydrocarbon" environmental samples. Orange points represent sampling locations for the enrichment cultures presented in this study.

Supplemental Table 1. Assembly statistics overview for *Nitrotoga* genomes. Percent completeness and contamination were estimated using CheckM. The asterisks indicate manual assignment of some marker genes used in completeness estimates (see Supplemental Note).

Supplemental Figure S1. Pairwise BLAST comparisons of 60 *Nitrotoga* and *Nitrotoga*-like 16S rRNA gene sequences ≥1,300 bp presented as a heatmap. The phylogeny from Figure 3 is displayed on the left with branch lengths ignored. Bolded sequence names have been enriched in culture. Nodes with a star represent organisms presented in this study.

Supplemental Figure S2. Partial alignment of important residues from selected NxrB subunits representing the *Nitrotoga* (Red), as well as canonical periplasmic-facing (Purple) and cytoplasmic-facing (Blue/Green) NXR. Highlights represent Fe-S cluster binding residues (Yellow).

Supplemental Figure S3. Schematic of electron transfer in *Nitrotoga* based on genomic evidence. **A)** Canonical nitrite oxidation performed by NXR will liberate two electrons onto cytochrome *c* and flow forwards to the terminal oxidase (Complex IV), or backwards to regenerate NADH via Alternative Complex III and the quinone pool, or to generate biochemical intermediates via Complex II. The question mark located in the NxrC subunit symbolizes the uncertainty in whether or not the holoenzyme is anchored to the cell membrane. **B)** Aerobic respiration with alternative electron donors such as NADH derived from organic carbon utilization. **C)** Hypothesized anaerobic respiration with alternative electron donors (e.g., NADH) with the reduction of nitrate to nitrite via NXR as seen for other NOB. **D)** Electrons derived from

reduced sulfur compounds (sulfites or sulfides) are transferred to cytochrome *c* or quinone, which can enter at any point in electron transfer shown in panels B or C.

Supplemental Figure S4. Relative abundance of **A)** *Nitrotoga*- and **B)** *Nitrospira*-like 16S rRNA gene sequence OTUs from water column ("WC") and sediment ("SED") samples in Bear Creek ("BEAR"), Cherry Creek ("CHERRY"), and the upstream ("SP-UP") and downstream ("SP-DOWN") sites at their respective confluences with the South Platte River. OTUs were based on the amplification with general 16S rRNA gene primers targeting the total bacterial community and were grouped at 97% nucleotide identity. *Nitrotoga* and *Nitrospira* OTUs were identified based on BLAST searches against the SILVA rRNA gene database. **C)** Log ratio of the summed relative abundance of *Nitrotoga*- to *Nitrospira*-like 16S rRNA gene sequences.

Supplemental File 1. List of accessions for 16S rRNA genes used in Figure 3 and Supplemental Figure S1.

Supplemental File 2. List of accessions for Type II DMSO reductase proteins used in Figure 4.

Unexpected versatility in the metabolism and ecophysiology of globally relevant nitrite-oxidizing *Nitrotoga* bacteria

Andrew M. Boddicker and Annika C. Mosier

SUPPLEMTENAL NOTE

METHODS

Culture inoculation and growth

Freshwater Nitrite Oxidizer Medium (FNOM) was prepared by mixing 1 g NaCl, 0.4 g $MgCl_2$-$6H_2O$, 0.1 g $CaCl_2$-$2H_2O$, 0.5 g KCl, 100 µL 10X vitamin solution (Balch *et al*., 1979), 1 mL 1M $NaHCO_3$, and 300 µL 1M $NaNO_2$ per liter. The pH of the media was lowered to 7.0 using 10% HCl, then autoclaved. After autoclaving, 10 mL of separately autoclaved 4 g $*$ L^{-1} KH_2PO_4 and 1 mL trace metal solution (Biebl and Pfennig, 1978) were sterilely added to the media before storing at 4°C in the dark.

Two surface sediment samples were aseptically collected in February 2015 from the urban-impacted Cherry Creek in downtown Denver, CO (samples MKT and LAW) using a cut-off sterile 30 mL syringe and returned to the lab on ice. The same day, the top 0.5 cm of sediment was mixed with 10 mL sterile FNOM, then 1 mL of the sediment slurry was transferred to 100 mL FNOM for incubation at room temperature in the dark. Two water column samples were collected in May 2015 from two agriculturally-impacted rivers near Greeley, CO (about 100 km North of Denver, CO) (samples CP45 from the Cache La Poudre River and SPKER from the South Platte River). River water was kept on ice in the field and stored at 4°C upon return to the lab. After five days, 10 mL of each water sample was transferred to 100 mL FNOM and allowed to incubate at room temperature in the dark.

Nitrite consumption was regularly monitored in the cultures using a Griess nitrite color reagent (Griess-Romijn van Eck, 1966) composed of 0.5 g sulfanilamide, 0.05 g N-(1-naphthyl) ethylenediamine dihydrochloride, 5 mL 85% phosphoric acid, and MilliQ water to a final volume of 50 mL. Nitrite color reagent was mixed with the culture at a 1:10 ratio for visual estimates (+/-) or at a 1:1 ratio for quantitative spectrophotometric measurements.

To determine rates of nitrite oxidation, 100 µL of each sample line (CP45, LAW, MKT, SPKER) was inoculated into three bottles with 100 mL FNOM. At regular intervals, samples were collected in triplicate from each bottle and mixed with equal volumes of fresh Griess nitrite color reagent, then the optical density (OD) was measured at 540, 545, and 550 nm using a BioTek Synergy HT plate reader (BioTek, Winooski, VT). The mean maximum OD was used to calculate nitrite concentrations based on a standard curve of sterile media ranging from 0-0.3 mM nitrite with the Gen5 analysis software (BioTek, Winooski, VT). Sterile FNOM was used as a negative control. Nitrite oxidation rates were calculated across three time points (R^2 value of 0.93-1.0) within logarithmic nitrite consumption for each bottle.

DNA extraction

At mid- to late-phase of exponential nitrite oxidation, 400 mL of each culture was filtered onto a 0.2 µm Supor 200 filter (Pall, New York, NY). Filters were cut into small pieces with a sterile scalpel and aseptically placed into a Lysing Matrix E Bead Beating Tube (MP Biomedicals, Santa Ana, CA) with 800 µL lysing buffer (750 mM sucrose, 20 mM EDTA, 400 mM NaCl, 50 mM Tris (pH 8.4)) and 100 µL 10% SDS. Samples were vortexed briefly before

bead beating in a FastPrep-24 5G reciprocating homogenizer (MP Biomedicals, Santa Ana, CA) at 5 m/s for 30 seconds. Samples were incubated at 99°C for 1-3 minutes before adding 50 µL 20 mg/mL proteinase K, then incubated for 3-5.5 hours in a rotating hybridization oven at 55°C. Cold 100% ethanol (500 µL) was added to each sample in the same tube, and then DNA was purified using the DNeasy Blood and Tissue Kit (following manufacturer's instructions for purification) (Qiagen, Hilden, Germany). Extracted DNA was quantified with a Qubit fluorometer (Thermo Fisher, Waltham, MA), using the High Sensitivity dsDNA assay.

Metagenome and *Nitrotoga* genome assembly

BBDuk v36.99 (http://jgi.doe.gov/data-and-tools/bbtools) was used to remove sequencing adapters and trim metagenomic reads (mink=8, hdist=1, qtrim=20, minlength=50, ftm=5, tpe, tbo). Read quality distributions were checked before and after trimming using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Metagenomes were assembled from each culture using the filtered and trimmed reads with MEGAHIT v1.0.6.1 (k-min=32, k-max=121, k-step=10). Reads were mapped to the metagenome assemblies using BBMap v36.x (http://jgi.doe.gov/data-and-tools/bbtools/) and contigs were binned using MetaBAT v.0.32.4 (--verysensitive -B20 --unbinned) (Kang *et al.*, 2015). Preliminary taxonomy of each genomic bin was identified based on (1) BLAST searches of 16S rRNA gene sequences (assembled using EMIRGE (Miller *et al.*, 2011)) against the SILVA 16S rRNA gene database (release 123) (Quast *et al.*, 2013); and (2) the CheckM lineage workflow (Parks *et al.*, 2015). In each culture, only one putative NOB was identified belonging to the *Nitrotoga* genus.

Putative *Nitrotoga* bins of interest were manually refined using the Anvi'o metagenomics pipeline version 2.1.0 (Eren *et al.*, 2015). Putative *Nitrotoga* bins were combined in the CP45 and LAW metagenomes as these genomes were each split into two different bins. The CheckM merge function (Parks *et al.*, 2015) was used to supervise bin mergers and the lineage workflow was run again after merging to ensure the completeness estimates increased and contamination estimates did not. To check for contaminant and chimeric contigs, genes were called with Prodigal (Hyatt *et al.*, 2010) and BLASTP (Camacho *et al.*, 2009) was used to find the best hit for each predicted protein against the UniRef90 database (release 2016_11) (Suzek *et al.*, 2014). Contigs with suspicious BLASTP taxonomy results were further scrutinized and removed upon later reassembly as needed.

For an iterative reassembly process of *Nitrotoga* genomes, reads were mapped to contigs of individual bins using BBSplit v36.x (http://jgi.doe.gov/data-and-tools/bbtools/) and were assembled using SPAdes v3.9.0 (Bankevich *et al.*, 2012) under the "careful" setting with MEGAHIT-assembled contigs given as "trusted contigs" and the same kmer range as used in metagenome assembly. Assembly graphs were visualized using Bandage (Wick *et al.*, 2015) to identify suspicious contigs. Contigs were removed from the bin before further reassembly if they had dissimilar best UniRef90 taxonomy hits compared to the rest of the contigs, inconsistent BLASTX taxonomy hits against the NCBI nr database, and were not found to be present within bins of other *Nitrotoga* genomes (from this study).

16S rRNA gene sequences (which rarely bin properly in metagenomic assemblies) were manually added to the *Nitrotoga* genome assemblies. To avoid selection bias, all 'unbinned' MEGAHIT-assembled contigs were searched against the SILVA 16S rRNA database (release 128) (Quast *et al.*, 2013) using BLASTN (Camacho *et al.*, 2009). All contigs with an alignment ≥300 bp (≥189 for the LAW metagenome due to a *Nitrotoga* hit of that size) to any 16S rRNA gene from the database were added to the *Nitrotoga* genome assemblies. The resulting assembly

graph was visualized in Bandage (Wick *et al.*, 2015), and the internal BLAST function was used to search for all added 16S rRNA suspected contigs. In each of the four *Nitrotoga* genomes, only one 16S rRNA gene assembled on a contig with paired reads mapped to other *Nitrotoga* genomic contigs. The assembled 16S rRNA genes from each culture were most similar to *Nitrotoga* sequences in SILVA. The correct contig containing the *Nitrotoga* 16S rRNA gene was kept, and others removed from the assembly.

A similar process was followed for adding contigs with nitrite oxidoreductase (*nxr*) gene sequences to the assembly. Predicted protein sequences from 'unbinned' MEGAHIT-assembled contigs were searched for members of the Type II DMSO reductase family (TIGR03479, TIGR03478, TIGR03477, TIGR03482) using HMMER3 (Eddy, 2011). All contigs with hits were added to the SPAdes reassembly if they had a coverage estimate that was similar to, or higher than, that of the respective *Nitrotoga* genome. Ultimately, a single unbinned contig in each of the CP45, LAW, and MKT metagenomes held all *nxr* genes. Eight unbinned contigs from the SPKER metagenome were used to assemble the *nxr* contig in the SPKER *Nitrotoga* genome.

Analysis of single nucleotide variants (SNVs) in Anvi'o (Eren *et al.*, 2015) indicated 32 SNVs across the SPKER *nxrA* gene, which are likely the result of variations between the three predicted *nxrA* copies in the SPKER *Nitrotoga* genome, and/or contaminant reads from one or more of the suspected *nxr* MEGAHIT-assembled contigs. Four of the SNVs were found in ~33% of mapped reads, indicating they may represent variants on one of the three suspected *nxrA* gene copies and caused contig breakage during assembly. None of the other *Nitrotoga nxrA* genes had SNVs and the SPKER *Nitrotoga nxrBCD* genes were intact with no SNVs. A consensus *nxrA* sequence is presented in this study for the SPKER *Nitrotoga* genome, as paired reads could not resolve the SNVs into individual gene copies.

Relative abundance of *Nitrotoga* in creek sediments and water column

Sediment samples from Bear Creek, Cherry Creek, and the South Platte River were collected using a sterile spatula and sterile petri dish. Prior to sampling at each site, the spatula was rinsed with 70% ethanol and wiped dry with a clean Kimwipe. The spatula was then rinsed with sterile water to remove residual ethanol, followed by site water. The bottom of the sterile petri dish was then placed down into the sediment, open-side-down. The sterile spatula was slid under the dish, trapping the sediment in the petri dish. The lid was put on the petri dish and wrapped in parafilm. All sediment samples were stored on dry ice for a maximum of three hours before arriving at the lab for permanent storage at -80°C.

Water samples were collected using sterile 1 L Nalgene bottles submerged approximately 15 cm in creek water undergoing constant flow. Nalgene bottles were rinsed with site water three times before sample acquisition. Following sample collection, samples were immediately placed on ice and transferred to the lab for processing for filtration within three hours. Water column samples were filtered onto 25 mm diameter, 0.22 µm pore size Supor membrane (Pall Corporation, Ann Arbor, MI) filters in a Swinnex (Merck Millipore, Burlington, MA) filter housing using a peristaltic pump (Series II Geopump, Geotech Environmental Equipment, Denver, CO). Following sample filtration, filters were stored at -80°C until DNA extraction. Pump tubing was sterilized prior to filtering each sample by sequential rinses of the pump tubing: (1) rinsed with 250 mL of autoclaved MilliQ water; (2) recirculation of 500 mL of 10% HCl through the tubing for three minutes; (3) rinsed again with 250 mL of autoclaved MilliQ water; and (4) rinsed with 250 mL of site water.

DNA was extracted using the MP Biomedicals FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA) according to kit instructions with homogenization in the FastPrep-24 5G reciprocating homogenizer at 6.0 m/sec for 40 seconds. The Qubit dsDNA HS and dsDNA BR Assay kits (Life Technologies, Carlsbad, CA) were used to determine the DNA concentration of extracts. Qubit DNA quantitation was run in duplicate to determine the average DNA concentration for each DNA extract.

Sequence processing was conducted using QIIME (Caporaso *et al.*, 2010). Paired-end reads were joined using fastq-join and filtered to a Phred quality score of 20. Sequences with < 80 bp merge length were discarded. The last 20 bp was removed from both ends of sequences after merging to remove primers. Processed reads were clustered into operation taxonomic units (OTUs) at 97% sequence identity and the DECIPHER web tool was used to check for chimeras (short sequences) (Wright *et al.*, 2012). Putative chimeras as well as OTUs with relative abundance < 0.05% were removed. All chloroplast OTUs and OTUs found in sequencing controls were also removed before analysis. Final taxonomy was assigned using a BLAST search against the SILVA 16S rRNA database (release 128) (Quast *et al.*, 2013).

RESULTS AND DISCUSSION

Nitrotoga genome assembly

CheckM indicated that the *Nitrotoga* genomes were near-complete based on a collection of 419 single-copy gene markers conserved within the Betaproteobacteria (UID3959). The LAW, MKT, and SPKER genomes were predicted to be 98.2% complete, while the CP45 genome was 97.0% complete due to the loss of three markers that were present on small contigs (<2 kb) removed before annotation (PF00731.15 AIR carboxylase; PF01259.13 Phosphoribosylaminoimidazolesuccinocarboxamide synthase; and TIGR02392 alternative sigma factor RpoH). All genomes were missing the same five marker genes (TIGR01745 aspartate-semialdehyde dehydrogenase; TIGR01574 tRNA-i(6)A37 thiotransferase enzyme MiaB; PF03618.9 Kinase-PPPase; TIGR01161 phosphoribosylaminoimidazole carboxylase, ATPase subunit; and PF13603.1 Leucyl-tRNA synthetase, Domain 2). A manual search for each marker using HMMER3 (Eddy, 2011) revealed strong (evalue <1e-41) hits to four of the five missing genes in all genomes. The identification of these four markers improved completeness estimates to 99.8% complete. The fifth marker (PF03618.9) had no hits from the *Nitrotoga* predicted protein sequences and was not found on any contig with suitable coverage among the 'unbinned' contigs. Homologs to this marker gene are found in the closest sequenced relatives: *Sideroxydans lithotrophicus* ES-1, *Gallionella capsiferriformans* ES-2, and *Gallionella acididurans* ShG14-8.

Nitrotoga genomes were estimated to contain 0.24-0.3% contamination. Specifically, the CP45, LAW, MKT genomes had duplications of the same two markers (PF09976.4 Tetratricopeptide repeat-like domain; and PF08340.6 domain of unknown function 1732), while the SPKER genome only had a duplicate of PF08340.6. Secondary copies had a drastically reduced HMM hit (evalue 2-11 orders of magnitude higher), so they may represent false-positives. None of the published Gallionellaceae genomes showed duplications of these genes.

Nitrotoga nitrogen metabolism

Nitrite oxidation

NxrA contains a specific nitrite/nitrate substrate binding channel, a molybdenum-bis(pyranopterin guanosine dinucleotide) (Mo-bisPGD) motif, and one iron-sulfur cluster for the conduction of electrons to the beta subunit (Daims, Lücker, & Wagner, 2016; Grimaldi, *et al*. 2013; Hille, Hall, & Basu, 2014; Magalon, *et al*. 2011). The beta subunit acts as an electron conductor, passing through three [4Fe-4S] clusters and one [3Fe-4S] cluster (Daims, Lücker, & Wagner, 2016; Grimaldi *et al*., 2013; Hille, Hall, & Basu, 2014; Magalon, *et al*. 2011). Finally, the variable gamma subunit, NxrC, is thought to bind 1-2 heme groups to transfer electrons to a cytochrome *c*, and is likely bound to the membrane, anchoring the NXR holoenzyme (Daims *et al.*, 2016). NOB genomes with periplasmic-facing NXR typically encode multiple candidate *nxrC* genes, with varying sizes and heme-binding components (Daims *et al*., 2015; Lücker *et al*., 2010; Lücker, *et al*., 2013; van Kessel *et al*., 2015)

Nitrotoga nxr genes were ultimately placed on single contigs within each genome forming an *nxrABCD* operon. *Nitrotoga* NxrA (TIGR03479, evalue < 2.2e-74) was 1169 amino acids in length in the CP45, MKT, and LAW genomes and 1155 aa in the SPKER genome due to a truncated N-terminus (Figure 4b). *Nitrotoga* NxrB (TIGR03478, evalue <2.6e-73) was 385 amino acids long in all genomes, while *Nitrotoga* NxrC (TIGR03477, evalue <2.8e-36) was 371 or 372 amino acids in length. Each contig also contained an NxrD delta subunit (TIGR03482, evalue <8.8e-26), which may act as a chaperone similar to TorD used in molybdenum cofactor assembly and protein folding (Hille *et al.*, 2014; Lücker *et al.*, 2013; Magalon *et al.*, 2011; Ngugi *et al.*, 2015).

Two conserved hypothetical proteins were found downstream of the *nxr* operon in most *Nitrotoga* genomes: a 70 amino acid protein of unknown function, and a 341 amino acid protein with conserved domains related to iron-transfer P-loop NTPases, which are required for cytosolic Fe-S cluster assembly (factor NBP35). These two genes are located on the same contig as the *nxr* genes in the CP45, LAW, and MKT genomes. The SPKER *nxr* contig contained only the *nxrABCD* operon, but the 341 amino acid protein was found on a neighboring contig. The CP45 *nxr* contig had additional truncated hypothetical proteins on either end of the contig.

Nitrotoga NXR subunits have a conserved protein structure that is highly divergent from other NXRs and members of the Type II DMSO reductase family of enzymes (Figure 4). *Nitrotoga* NxrA subunits were at least 98.1% identical to each other across the 1169 amino acid protein but were as low as 84.1% identical across the nucleotide alignment (SPKER vs LAW). Similar patterns were seen with the beta subunit (≥97.9% amino acid identity; ≥87.0% nucleotide identity), while the gamma subunit was slightly more divergent (≥92.8% amino acid identity; ≥85.3% nucleotide identity). The delta chaperone was also highly conserved between *Nitrotoga* genomes (≥96.4% amino acid identity; ≥87.6% nucleotide identity).

Dissimilatory and assimilatory nitrogen metabolism

Urea is an important source of ammonia for many NOB (e.g., *Nitrospina* and some *Nitrospira*) (Daims *et al.*, 2016; Ushiki *et al.*, 2018). A urea-binding-like protein was found in the LAW genome located near most of the nitrogen transport systems (i.e., *narK* nitrite/nitrate transporter, nitrogen metabolism transcription factor *ntrC*, formate/nitrite transporter, *nirK* assimilatory nitrite reductase, and ammonia transporter *amtB*). However, no urea transporter or urease genes were identified in any *Nitrotoga* genomes.

Each *Nitrotoga* genome contained a nitrilase enzyme that degrades nitriles (C≡N bonds) to ammonia and carboxylic acids. Nitrilase could potentially be used as defense against simple nitriles like cyanide or to cleave ammonia for assimilation within the *Nitrotoga* cells. Previously, a cyanate-degrading NOB (via a cyanase enzyme) was shown to participate in 'reciprocal feeding' by releasing ammonia for consumption by ammonia-oxidizers (Palatinszky *et al.*, 2015). However, no cyanate transporter or cyanase genes were identified in any *Nitrotoga* genomes. Each *Nitrotoga* genome did contain four unique genes with rhodanese domains, which are known to detoxify cyanide (CN^-) using thiosulfate ($S_2O_3^{2-}$) as a sulfur donor, producing thiocyanate (SCN^-) and sulfite (SO_3^{2-}). One of these genes was predicted to be anchored into the cell membrane and face the periplasm. Thus, *Nitrotoga* may be capable of detoxifying cyanide while also contributing to their sulfur metabolism via sulfite production (see *sulfur energetics* section).

Nitrotoga carbon metabolism

Carbon fixation

The CP45, LAW, and MKT genomes had two copies of both small and large ribulose 1,5-bisphosphate carboxylase (RuBisCO), the key enzyme for CO_2 fixation, while the SPKER genome had single copies. All *Nitrotoga* genomes had an active form IC/ID RuBisCO, while the second copy in the CP45, LAW, and MKT genomes was related to form IA RuBisCO. All form I RuBisCO are likely active in carbon fixation (unlike most Form IV RuBisCO (Tabita *et al.*, 2007)) and subtypes IA, IC, and ID have been found across the Proteobacteria (Tabita *et al.*, 2008).

The two key enzymes of the reverse tricarboxylic acid (rTCA) cycle (oxoglutarate:ferredoxin oxidoreductase and ATP-citrate lyase) were missing from *Nitrotoga* genomes. The rTCA cycle is used in *Nitrospira*, *Nitrospina*, and *Ca.* Nitromaritima species to fix carbon dioxide (Lücker *et al.*, 2010, 2013; Ngugi *et al.*, 2015), while the *Nitrolancea*, *Nitrococcus*, and *Nitrobacter* utilize the Calvin cycle (Sorokin *et al.*, 2012; Füssel *et al.*, 2017; Starkenburg *et al.*, 2008).

Polysaccharide storage and Exopolysaccharide

Glycogen and starch storage is likely in *Nitrotoga* given the presence of key genes (i.e., glucose-1-phosphate adenylyltransferase, 1,4-alpha-glucan branching enzymes, alpha-amylase, starch synthase, and starch phosphorylase). A pathway for cellulose production (cellulose synthase (UDP-forming)) and hydrolysis (cellulase/endogluconase and cellobiose phosphorylase) to glucose 1-phosphate was also present in *Nitrotoga* genomes, except the LAW genome was missing a cellobiose phosphorylase gene.

Exopolysaccharide (EPS) synthesis is also likely in *Nitrotoga*, as an extensive array of PEP-CTERM exopolysaccharide sorting enzymes (Haft *et al.*, 2006), polysaccharide export pumps, and saccharide modifiers (i.e., UDP-glucose dehydrogenase, polyprenyl glycosylphosphotransferase, D-alanyl-lipoteichoic acid acyltransferase) were present in all genomes. EPS production has been observed in *Nitrotoga* via microscopy (Ishii *et al.*, 2017), and may explain the formation of microcolonies in WWTPs (Lücker *et al.*, 2015).

Nitrotoga iron acquisition

The CP45, LAW, and SPKER genomes included two genes with sequence similarity to IucA/IucC and FhuF domains that are commonly found in nonribosomal peptide synthase-

independent siderophore biosynthesis (Challis, 2005). However, it is unclear whether *Nitrotoga* are capable of complete siderophore biosynthesis as these genes also resemble a ferric reductase, which may be used to reduce siderophore-bound iron prior to transport across the cell membrane. Reduced Fe^{2+} is likely transported into the cytoplasm via an EfeU Iron/Lead type transporter, while an FhuDBC ABC-type transporter can transport some complete siderophores across the cell membrane. An AfuABC ABC-type transporter was also found with predicted Fe^{3+} transporter activity, however recent evidence suggests this may actually be a phosphorylated carbohydrate pump (Sit *et al.*, 2015).

Fe^{3+} storage is possible with bacterioferritin (BfrB) and a bacterioferritin-associated ferredoxin (Bfd) encoded in *Nitrotoga* genomes. However, a ferredoxin NADP reductase (Fpr) responsible for transferring electrons to Bfd, and ultimately Fe^{3+} for mobilization into the cytoplasm (Rivera, 2017; Wang *et al.*, 2015), was not detected in *Nitrotoga* genomes. A short ferritin-like protein, also useful for iron storage, was encoded in the CP45, LAW, and SPKER genomes although its function is unknown. Fur family transcriptional regulators, which measure cytoplasmic iron concentrations, were found near the *afuABC* operon and a zinc/cadmium transporter in *Nitrotoga* genomes and may respond to changes in iron availability.

A heme-degrading monooxygenase gene, *hmoA*, used to harvest heme from a host, was found in all *Nitrotoga* genomes. The presence of two putative hemolysin genes in all *Nitrotoga* genomes, and an additional heme oxygenase in the CP45 genome, may support an antagonistic role of *Nitrotoga* in bacterial communities. However, heme binding and uptake systems were not observed, and only the LAW genome encodes a Type I protein secretion system.

REFERENCES

Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS. (1979). Methanogens: Reevaluation of a Uniqe Biological Group. *Microbiol Rev* **43**: 260–296.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**: 455–477.

Biebl H, Pfennig N. (1978). Growth yields of green sulfur bacteria in mixed cultures with sulfur and sulfate reducing bacteria. *Arch Microbiol* **117**: 9–16.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al.* (2009). BLAST plus: architecture and applications. *BMC Bioinformatics* **10**: 1.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Challis GL. (2005). A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *ChemBioChem* **6**: 601–611.

Daims H, Lebedeva E V., Pjevac P, Han P, Herbold C, Albertsen M, *et al.* (2015). Complete nitrification by Nitrospira bacteria. *Nature*. e-pub ahead of print, doi: 10.1038/nature16461.

Daims H, Lücker S, Wagner M. (2016). A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol* **24**: 699–712.

Eddy SR. (2011). Accelerated profile HMM searches. *PLoS Comput Biol* **7**. e-pub ahead of print, doi: 10.1371/journal.pcbi.1002195.

Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, *et al.* (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: e1319.

Füssel J, Lücker S, Yilmaz P, Nowka B, van Kessel MAHJ, Bourceau P, *et al.* (2017). Adaptability as the key to success for the ubiquitous marine nitrite oxidizer *Nitrococcus*. *Sci Adv* **3**: e1700807.

Griess-Romijn van Eck E. (1966). Physiological and chemical tests for drinking water. (Nederlands Normalisatie Instituut).

Grimaldi S, Schoepp-Cothenet B, Ceccaldi P, Guigliarelli B, Magalon A. (2013). The prokaryotic Mo/W-bisPGD enzymes family: A catalytic workhorse in bioenergetic. *Biochim Biophys Acta - Bioenerg* **1827**: 1048–1085.

Haft DH, Paulsen IT, Ward N, Selengut JD. (2006). Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic. *BMC Biol* **4**: 29.

Hille R, Hall J, Basu P. (2014). The Mononuclear Molybdenum Enzymes. *Chem Rev* **114**: 3963–4038.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Ishii K, Fujitani H, Soh K, Nakagawa T, Takahashi R, Tsuneda S. (2017). Enrichment and Physiological Characterization of a Cold-Adapted Nitrite Oxidizer Nitrotoga sp. from Eelgrass Sediments. *Appl Environ Microbiol* **83**: 1–14.

Kang DD, Froula J, Egan R, Wang Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.

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

Lücker S, Nowka B, Rattei T, Spieck E, Daims H. (2013). The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front Microbiol* **4**: 1–19.
Lücker S, Schwarz J, Gruber-Dorninger C, Spieck E, Wagner M, Daims H. (2015). Nitrotoga-like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants. *ISME J* **9**: 708–720.
Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B, *et al.* (2010). A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci U S A* **107**: 13479–13484.
Magalon A, Fedor JG, Walburger A, Weiner JH. (2011). Molybdenum enzymes in bacteria and their maturation. *Coord Chem Rev* **255**: 1159–1178.
Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF, Pace N, *et al.* (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **12**: R44.
Ngugi DK, Blom J, Stepanauskas R, Stingl U. (2015). Diversification and niche adaptations of Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *ISME J* **10**: 1383–1399.
Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, *et al.* (2015). Cyanate as an energy source for nitrifiers. *Nature* **524**: 105–108.
Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–55.
Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, *et al.* (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**: 590–596.
Rivera M. (2017). Bacterioferritin: Structure, Dynamics, and Protein-Protein Interactions at Play in Iron Storage and Mobilization. *Acc Chem Res* **50**: 331–340.
Sit B, Crowley SM, Bhullar K, Lai CCL, Tang C, Hooda Y, *et al.* (2015). Active Transport of Phosphorylated Carbohydrates Promotes Intestinal Colonization and Transmission of a Bacterial Pathogen. *PLoS Pathog* **11**: 1–22.
Sorokin DY, Lücker S, Vejmelkova D, Kostrikina N a, Kleerebezem R, Rijpstra WIC, *et al.* (2012). Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J* **6**: 2245–2256.
Starkenburg SR, Larimer FW, Stein LY, Klotz MG, Chain PSG, Sayavedra-Soto LA, *et al.* (2008). Complete genome sequence of Nitrobacter hamburgensis X14 and comparative genomic analysis of species within the genus Nitrobacter. *Appl Environ Microbiol* **74**: 2852–63.
Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. (2014). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**: 926–932.
Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. (2007). Function, Structure, and Evolution of the RubisCO-Like Proteins and Their RubisCO Homologs. *Microbiol Mol Biol Rev* **71**: 576–599.
Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* **59**: 1515–1524.
Ushiki N, Fujitani H, Shimada Y, Morohoshi T. (2018). Genomic Analysis of Two Phylogenetically Distinct Nitrospira Species Reveals Their Genomic Plasticity and Functional

Diversity. *Front Microbiol* **8**: 1–12.
Wang Y, Yao H, Cheng Y, Lovell S, Battaile KP, Midaugh CR, *et al.* (2015). Characterization of the Bacterioferritin/Bacterioferritin Associated Ferredoxin Protein-Protein Interaction in Solution and Determination of Binding Energy Hot Spots. *Biochemistry* **54**: 6162–6175.
Wick RR, Schultz MB, Zobel J, Holt KE. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **31**: 3350–3352.
Wright ES, Yilmaz LS, Noguera DR. (2012). DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* **78**: 717–725.

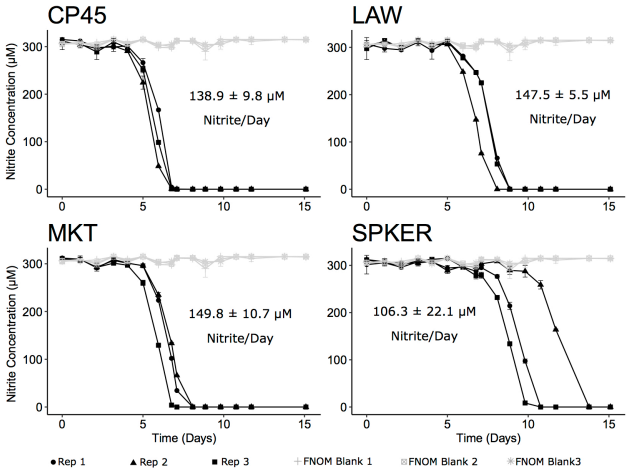


Figure 1. Nitrite consumption by *Nitrotoga* enrichment cultures over time. Each enrichment culture was inoculated in three replicates, and nitrite concentration was quantified colorimetrically in triplicate at each time point. Error bars show the standard deviation of each time point; error bars that appear to be missing are too small to be visualized. Sterile FNOM was used as a control and plotted with each culture. Logarithmic declines in nitrite concentration were used to calculate the nitrite oxidation rate for each biological replicate. The average nitrite consumption per day is shown with standard deviation among triplicate cultures.

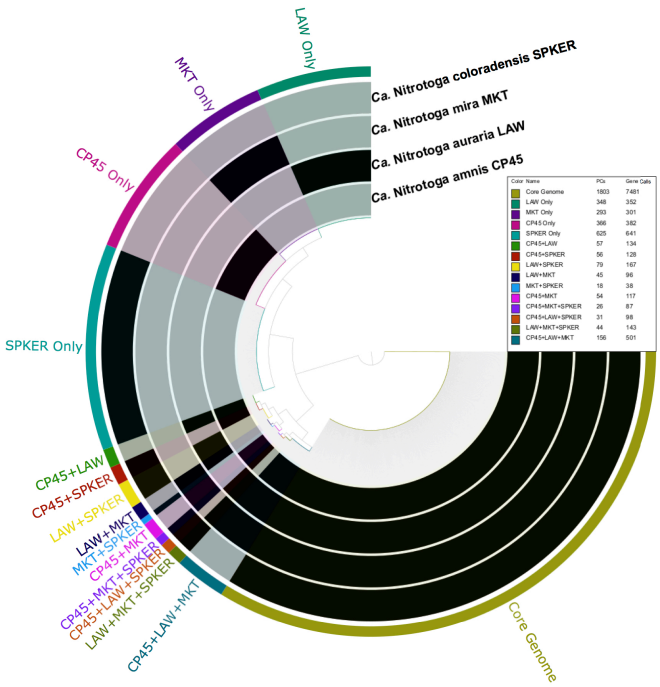
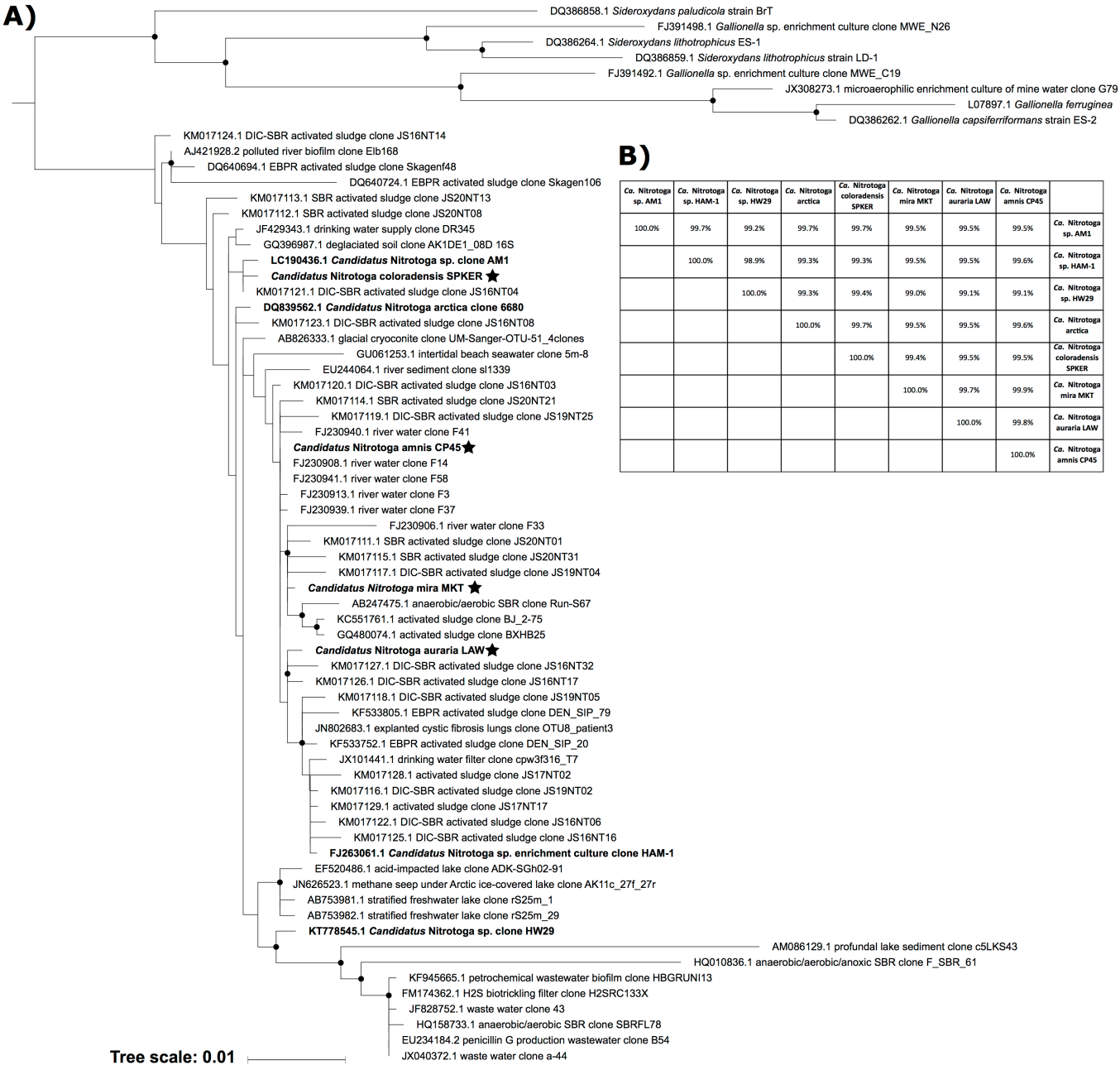


Figure 2. Anvi'o pangenome analysis of four *Nitrotoga* genomes. Coding sequences for all four genomes (10,666) in total grouped into 4,001 protein clusters (PCs), based on a pairwise BLAST of all coding sequences from all *Nitrotoga* genomes and clustering using the MCL algorithm (mcl=10). The core genome was represented by 1,803 PCs found in all four *Nitrotoga* genomes. Each individual genome (e.g., MKT only) contained 293-625 unique PCs not found in any other *Nitrotoga* genome.



B)

Ca. Nitrotoga sp. AM1	*Ca.* Nitrotoga sp. HAM-1	*Ca.* Nitrotoga sp. HW29	*Ca.* Nitrotoga arctica	*Ca.* Nitrotoga coloradensis SPKER	*Ca.* Nitrotoga mira MKT	*Ca.* Nitrotoga auraria LAW	*Ca.* Nitrotoga amnis CP45	
100.0%	99.7%	99.2%	99.7%	99.7%	99.5%	99.5%	99.5%	*Ca.* Nitrotoga sp. AM1
	100.0%	98.9%	99.3%	99.3%	99.5%	99.5%	99.6%	*Ca.* Nitrotoga sp. HAM-1
		100.0%	99.3%	99.4%	99.0%	99.1%	99.1%	*Ca.* Nitrotoga sp. HW29
			100.0%	99.7%	99.5%	99.5%	99.6%	*Ca.* Nitrotoga arctica
				100.0%	99.4%	99.5%	99.5%	*Ca.* Nitrotoga coloradensis SPKER
					100.0%	99.7%	99.9%	*Ca.* Nitrotoga mira MKT
						100.0%	99.8%	*Ca.* Nitrotoga auraria LAW
							100.0%	*Ca.* Nitrotoga amnis CP45

Figure 3. A) Maximum likelihood phylogenetic tree of 16S rRNA gene sequences from representative *Ca.* Nitrotoga sequences and close relatives. Sequences were aligned across 1,422 positions; phylogenetic trees were generated using RAxML (version 8.2.9) with 1,000 rapid bootstraps and the GTRGAMMA model of nucleotide substitution. Nodes with bootstrap support values ≥50% are shown with a black circle. A midpoint root was used when selecting outgroup sequences. Bolded sequence names have been enriched in culture. Nodes with a star represent organisms presented in this study. **B)** BLASTN comparisons of near full-length 16S rRNA gene sequences from the eight known enriched *Nitrotoga* sp.

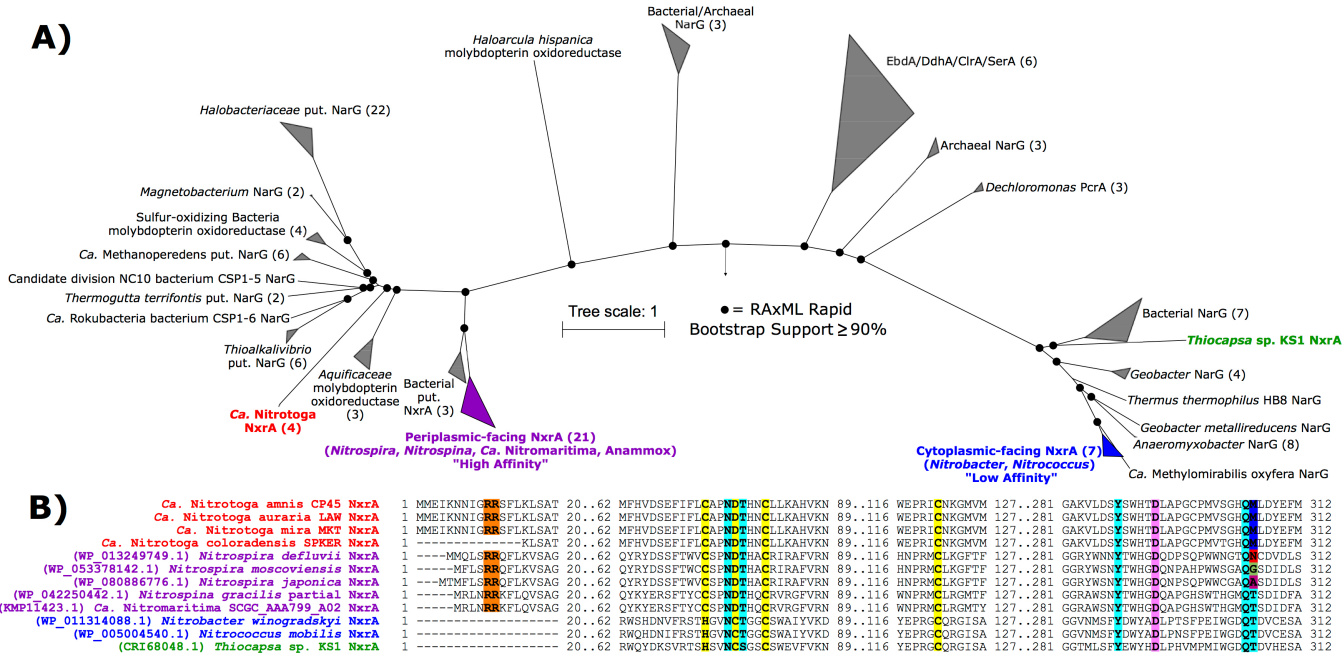
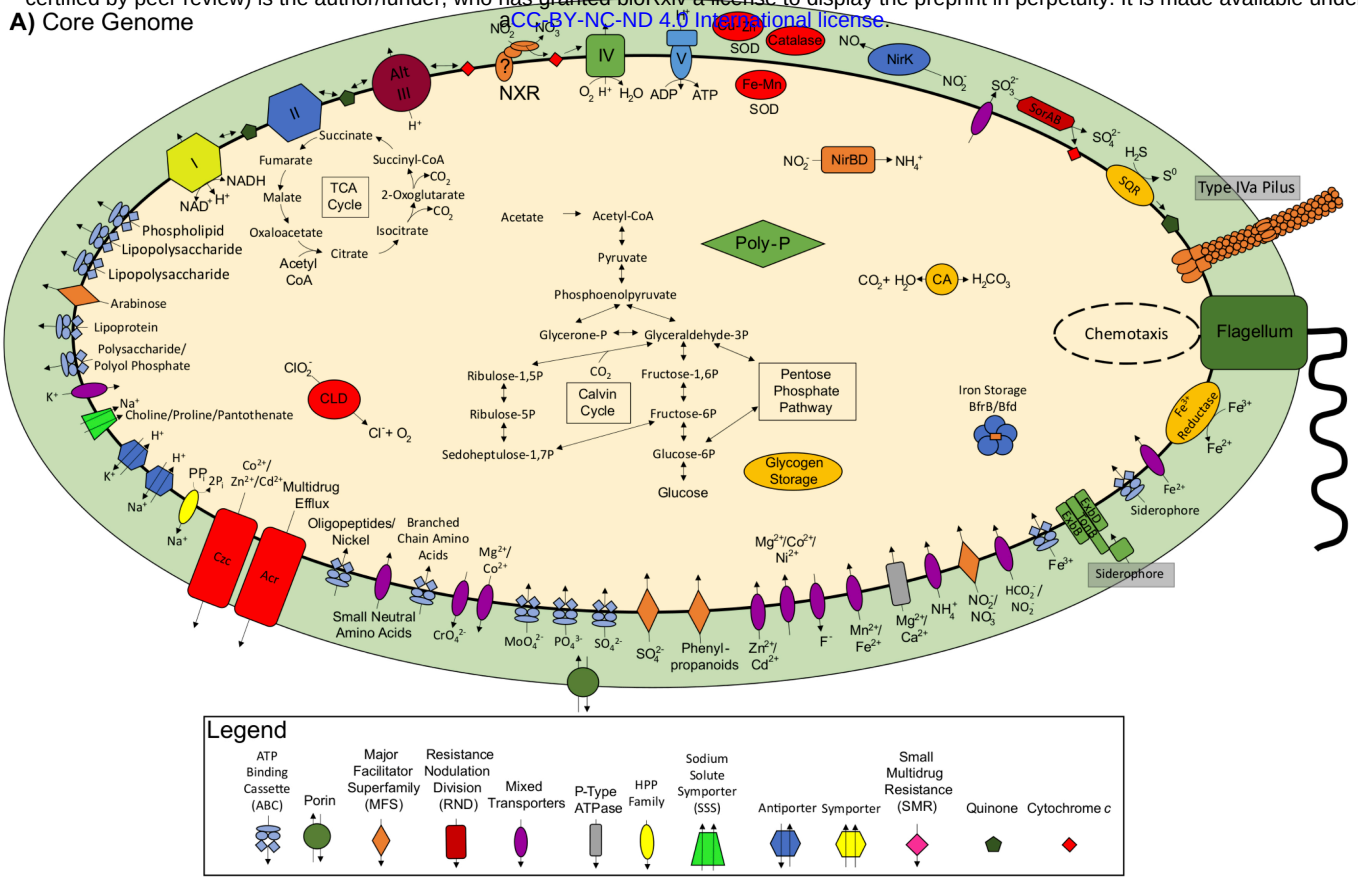


Figure 4. Phylogenetic and structural analysis of the alpha subunit of nitrite oxidoreductase (NxrA). **A)** Phylogeny of 122 members of the Type II DMSO reductase protein family aligned and manually trimmed to 1,651 amino acid positions. References were selected to include both cytoplasmic-facing "Low Affinity" and periplasmic-facing "High Affinity" NxrA (Ngugi *et al*., 2015), as well as other family members: PcrA (perchlorate reductase), EbdA (ethylbenzene dehydrogenase), DdhA (dimethylsulfide dehydrogenase), ClrA (chlorate reductase), SerA (selenite reductase), and NarG (nitrate reductase). Putative enzymes are marked as "put.". The number of sequences in collapsed nodes are shown in parentheses. RAxML rapid boostrap support values ≥90% are marked with a black circle. **B)** Partial alignment of selected important residues in *Nitrotoga*, periplasmic-facing, and cytoplasmic-facing NxrA (including *Thiocapsa*). Highlights represent Tat signal peptides (Orange), Fe-S cluster binding residues (Yellow), molybdenum coordinating residues (Pink), and nitrite/nitrate binding residues (Light Blue).



B) Accessory Genomes

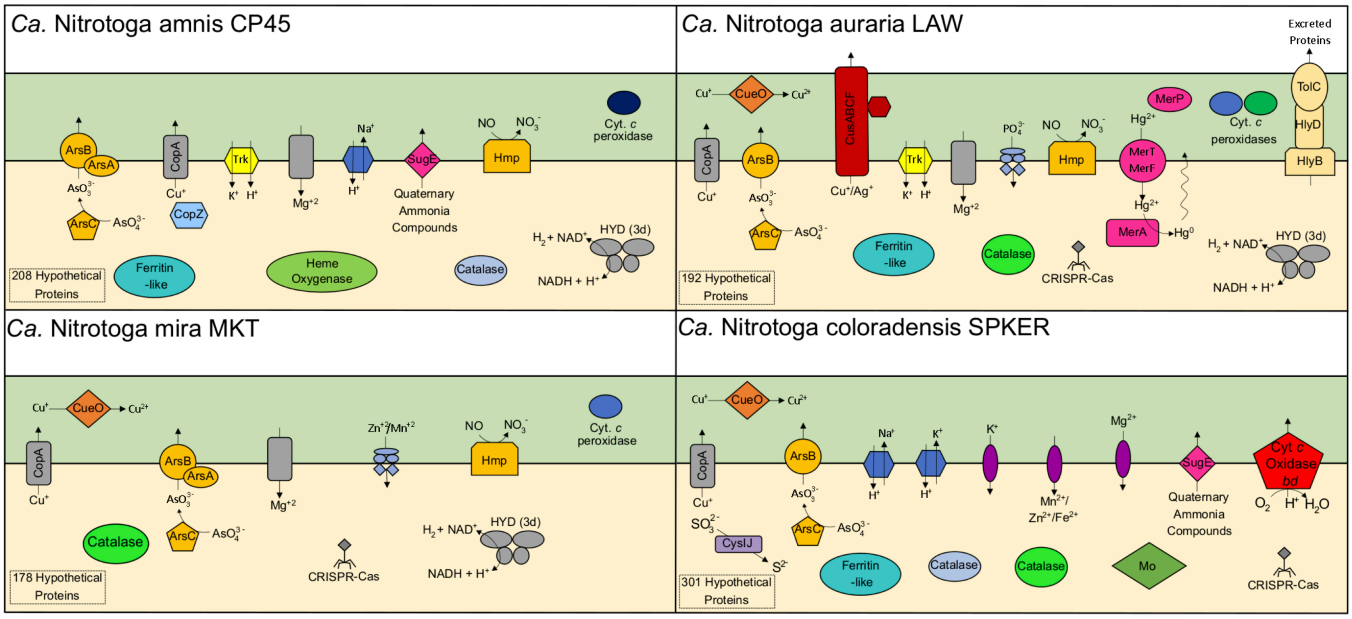


Figure 5. Schematic representation of *Nitrotoga* genome features. **A)** The core *Nitrotoga* genome including predicted functions shared by all four genomes. The question mark located in the NxrC subunit symbolizes the uncertainty in whether or not the holoenzyme is anchored to the cell membrane. **B)** Predicted function of individual genomic features found within one or more genomes, but not shared among all four. Each genome had 178-301 unique hypothetical proteins, and an additional 171 hypothetical proteins were shared among two or three genomes.

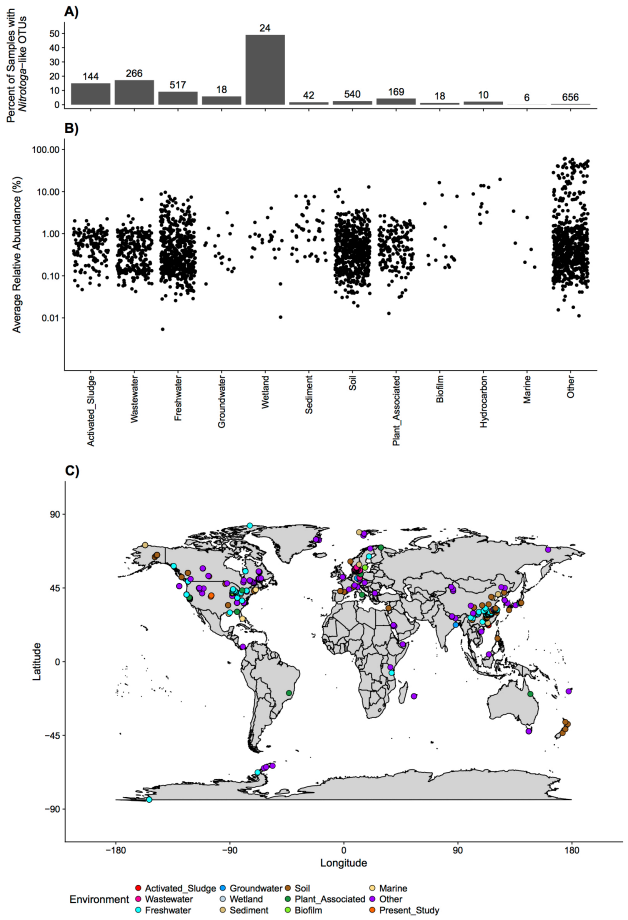
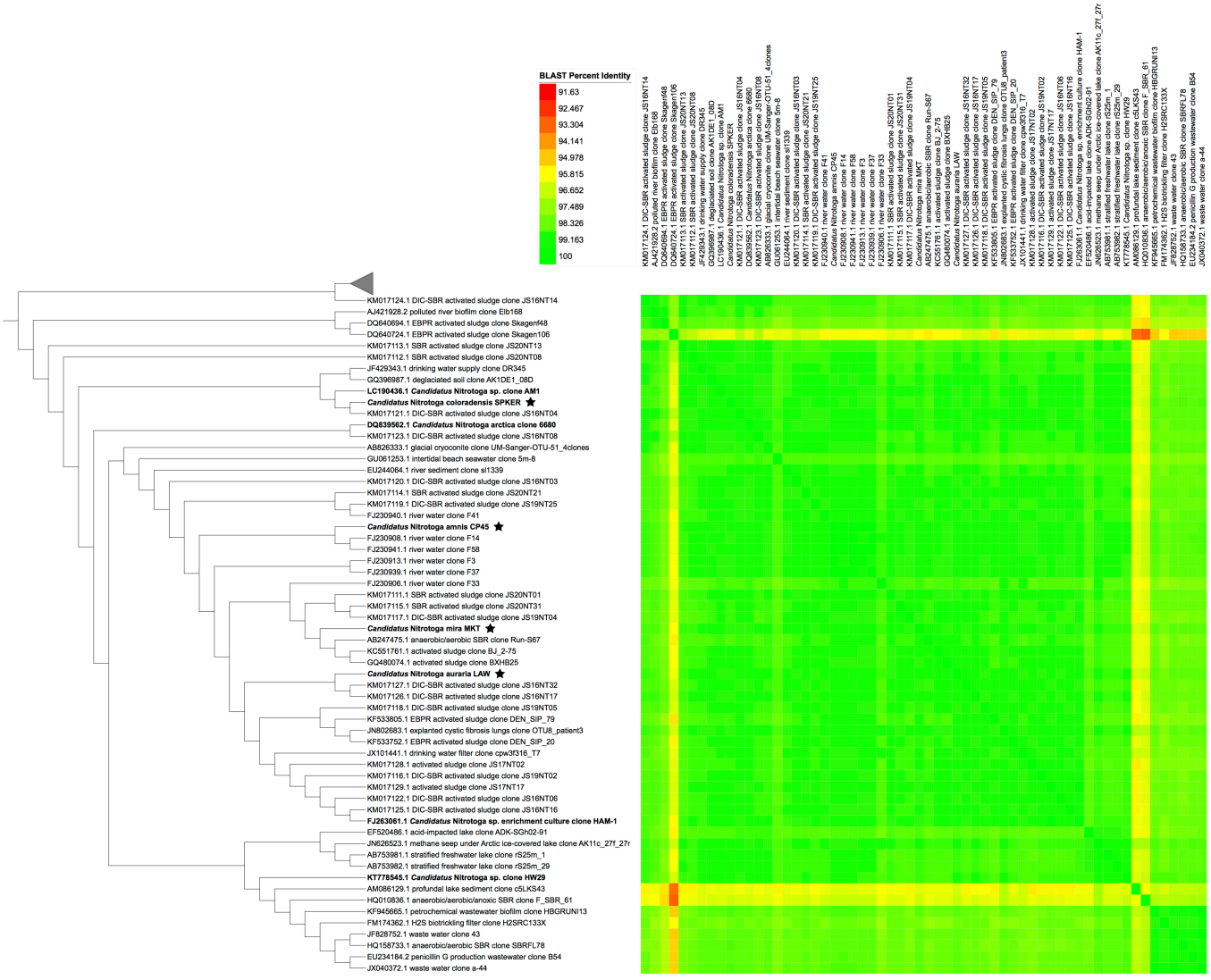
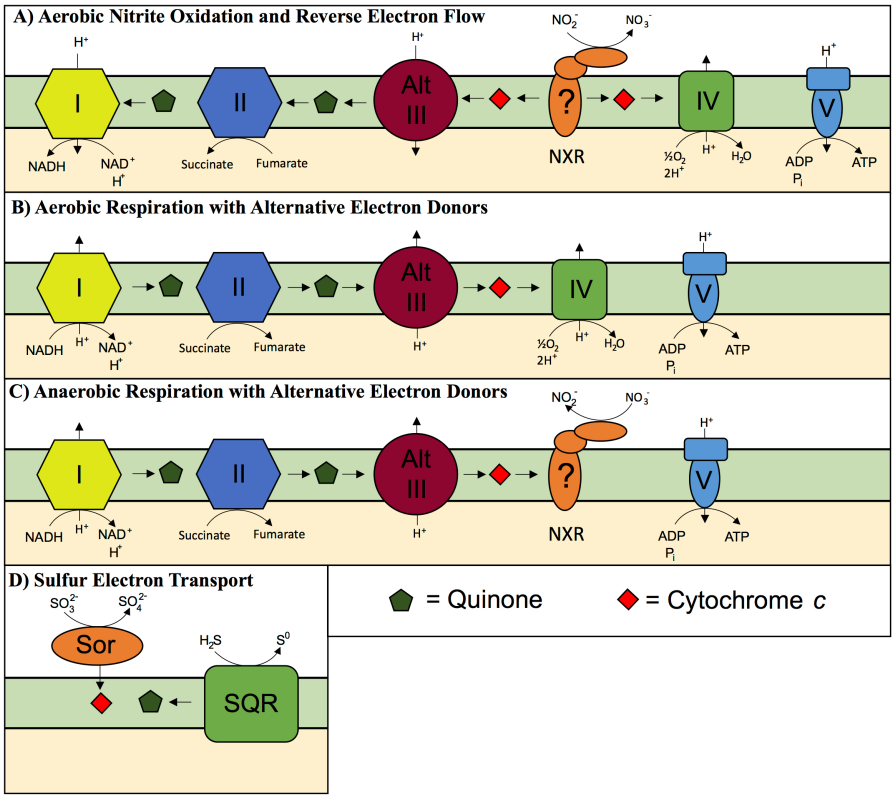


Figure 6. Global distribution of *Nitrotoga*-like sequences. OTUs from 16S rRNA gene amplicon studies deposited as SRA runs were clustered by IMNGS. All runs with OTUs ≥ 97% identity to cultured *Nitrotoga* 16S rRNA gene sequences (from this study), an alignment of at least 200 bp, and at least 100 total reads were kept. **A)** The percent of SRA runs with *Nitrotoga*-like OTUs within respective environments are shown with the number of SRA runs with *Nitrotoga*-like OTUs displayed above each bar. **B)** The relative abundance of *Nitrotoga*-like OTUs was averaged across all four queried *Nitrotoga* 16S rRNA gene sequences and plotted by environment. **C)** Global distribution of SRA runs with *Nitrotoga*-like OTUs. 484 of the 2,410 SRA runs with *Nitrotoga*-like OTUs did not have geographic information available, including all of the "Hydrocarbon" environmental samples. Orange points represent sampling locations for the enrichment cultures presented in this study.

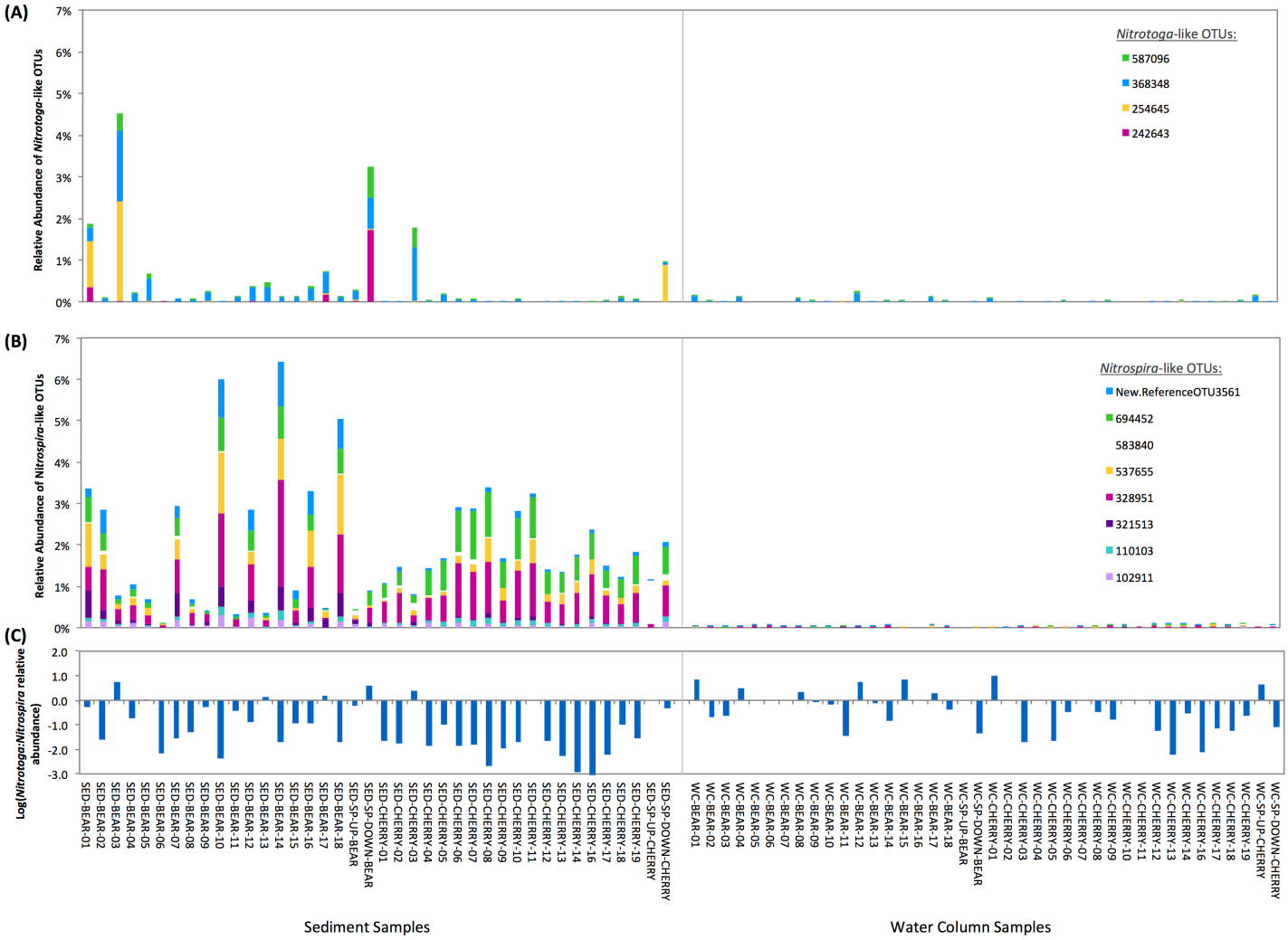


Supplemental Figure S1. Pairwise BLAST comparisons of 60 *Nitrotoga* and *Nitrotoga*-like 16S rRNA gene sequences ≥1,300 bp in length presented as a heatmap. The phylogeny from Figure 3 is displayed on the left with branch lengths ignored. Bolded sequence names have been enriched in culture. Nodes with a star represent organisms presented in this study.

Supplemental Figure S2. Partial alignment of important residues from selected NxrB subunits representing the *Nitrotoga* (Red), as well as canonical periplasmic-facing (Purple) and cytoplasmic-facing (Blue/Green) NXR. Highlights represent Fe-S cluster binding residues (Yellow).



Supplemental Figure S3. Schematic of electron transfer in *Nitrotoga* based on genomic evidence. **A)** Canonical nitrite oxidation performed by NXR will liberate two electrons onto cytochrome *c* and flow forwards to the terminal oxidase (Complex IV), or backwards to regenerate NADH via Alternative Complex III and the quinone pool, or to generate biochemical intermediates via Complex II. The question mark located in the NxrC subunit symbolizes the uncertainty in whether or not the holoenzyme is anchored to the cell membrane. **B)** Aerobic respiration with alternative electron donors such as NADH derived from organic carbon utilization. **C)** Hypothesized anaerobic respiration with alternative electron donors (e.g., NADH) with the reduction of nitrate to nitrite via NXR as seen for other NOB. **D)** Electrons derived from reduced sulfur compounds (sulfites or sulfides) are transferred to cytochrome *c* or quinone, which can enter at any point in electron transfer shown in panels B or C.



Supplemental Figure S4. Relative abundance of **A)** *Nitrotoga*- and **B)** *Nitrospira*-like 16S rRNA gene sequence OTUs from water column ("WC") and sediment ("SED") samples in Bear Creek ("BEAR"), Cherry Creek ("CHERRY"), and the upstream ("SP-UP") and downstream ("SP-DOWN") sites at their respective confluences with the South Platte River. OTUs were based on the amplification with general 16S rRNA gene primers targeting the total bacterial community and were grouped at 97% nucleotide identity. *Nitrotoga* and *Nitrospira* OTUs were identified based on BLAST searches against the SILVA rRNA gene database. **C)** Log ratio of the summed relative abundance of *Nitrotoga*- to *Nitrospira*-like 16S rRNA gene sequences.

#supplemental_file_1_16S_rRNA_gene_accessions
JF429343.1
GQ396987.1
KM017113.1
KM017124.1
DQ640724.1
DQ640694.1
AJ421928.2
AB826333.1
EF520486.1
AB753981.1
JN626523.1
AB753982.1
KT778545.1
AM086129.1
HQ010836.1
HQ158733.1
KF945665.1
JF828752.1
EU234184.2
JX040372.1
FM174362.1
KM017123.1
KM017112.1
KM017121.1
DQ839562.1
KM017127.1
KM017126.1
KF533805.1
KF533752.1
JN802683.1
KM017118.1
KM017128.1
KM017125.1
JX101441.1
KM017116.1
FJ263061.1
KM017122.1
KM017129.1
KM017111.1
FJ230913.1
FJ230939.1
KM017115.1
KM017120.1
KM017119.1
FJ230941.1
FJ230908.1
FJ230940.1
FJ230906.1
KM017114.1

EU244064.1
KM017117.1
GU061253.1
AB247475.1
KC551761.1
GQ480074.1
LC190436.1
L07897.1
DQ386262.1
DQ386264.1
DQ386858.1
DQ386859.1
FJ391492.1
FJ391498.1
JX308273.1

#supplemental_file_2_Type_II_DMSO_protein_accessions
WP_005557333.1
KRT71940.1
WP_012964344.1
OHB89378.1
ARP53735.1
WP_006671140.1
WP_005553455.1
WP_015320288.1
WP_090306927.1
WP_049927103.1
WP_008426537.1
WP_008895484.1
WP_012942830.1
WP_049923369.1
WP_013880708.1
WP_090615652.1
WP_007701576.1
WP_006652972.1
WP_071400145.1
WP_006166072.1
AGB33389.1
WP_006182857.1
WP_006649791.1
WP_007142113.1
WP_089784439.1
WP_014030951.1
WP_015910069.1
WP_006671327.1
WP_012459851.1
Q9S1H0.1
P60068.1
WP_012173623.1
WP_028989412.1
WP_004513191.1
KPQ43397.1
EQB62431.1
KRT68883.1
WP_012470696.1
WP_012662440.1
WP_005024327.1
CAF21906.1
WP_011509502.1
WP_011511827.1
WP_041359063.1
WP_011314088.1
WP_011315305.1
WP_004998773.1
WP_005004540.1
CBE67843.1

WP_012468292.1
WP_012132377.1
WP_000040373.1
WP_012017453.1
WP_011365654.1
WP_004512301.1
CAA71210.2
WP_010877688.1
YP_430751.1
CRI68048.1
WP_011344974.1
WP_011879804.1
WP_089723899.1
OQY55899.1
KCZ70344.1
WP_048094100.1
KMP11423.1
WP_048492943.1
WP_042250442.1
WP_042251421.1
WP_015257936.1
WP_018861559.1
WP_018868538.1
WP_018863956.1
WP_019624109.1
WP_040333786.1
KJU86102.1
KHE93157.1
CAJ72445.1
WP_052562588.1
WP_096601733.1
WP_012513384.1
WP_052491869.1
OQW90655.1
ACL65121.1
WP_012633045.1
ACG72937.1
WP_012525755.1
GAO02124.1
WP_059436536.1
ABC81940.1
WP_011421222.1
WP_097297472.1
SNQ62035.1
WP_096206654.1
WP_062483509.1
CUS33249.1
WP_053378142.1
WP_053381280.1
WP_053381686.1

WP_053381689.1
WP_053381277.1
WP_080886776.1
CUS31266.1
WP_080885591.1
WP_013249749.1
WP_013249767.1
CUS38776.1
WP_080885705.1
OHB90204.1
Q8GPG4.1
WP_011223493.1
AAO49008.1
AFE03185.1
Q47CW6.1
WP_018232354.1
ASV74111.1
WP_095414528.1
AAK76387.1

Supplementary Table 1. Assembly statistics overview for *Nitrotoga* genomes. Percent completeness and contamination were estimated using CheckM. The asterisks indicate manual assignment of some marker genes used in completeness estimates (see Supplemental Note).

	Num. Contigs	GC %	N50 (bp)	Median Contig Length (bp)	Total Length (bp)	CDS	tRNA	rRNA	% Complete*	% Contamination
Ca. Nitrotoga amnis CP45	59	48.8%	69,731	35,034	2,816,237	2,665	36	6	99.8%	0.30%
Ca. Nitrotoga auraria LAW	33	48.7%	123,222	58,971	2,814,534	2,730	37	6	99.8%	0.30%
Ca. Nitrotoga mira MKT	23	48.6%	214,834	70,433	2,707,194	2,574	39	6	99.8%	0.30%
Ca. Nitrotoga coloradensis SPKER	32	47.5%	168,712	48,253	2,982,257	2,858	39	6	99.8%	0.24%

Table 1. Average nucleotide identity (ANI) between enriched *Nitrotoga* genomes and close relative genomes available on NCBI (unhighlighted cells; top right) and average amino acid identity (AAI) pairwise comparisons (highlighted cells; bottom left).

	Ca. Nitrotoga amnis CP45	***Ca.* Nitrotoga auraria LAW**	***Ca.* Nitrotoga mira MKT**	***Ca.* Nitrotoga coloradensis SPKER**	***Sideroxydans lithotrophicus* ES-1**	***Gallionella capsiferrifor-mans* ES-2**	***Ca.* Gallionella acididurans ShG14-8**	***Ferriphaselus amnicola***	***Ferriphaselus* sp. R-1**
Ca. Nitrotoga amnis CP45	100.0%	92.6%	93.3%	86.2%	76.2%	76.2%	75.9%	74.5%	76.2%
***Ca.* Nitrotoga auraria LAW**	93.8%	100.0%	93.4%	86.2%	74.2%	74.3%	74.8%	73.8%	74.1%
***Ca.* Nitrotoga mira MKT**	93.9%	94.6%	100.0%	85.9%	75.6%	76.5%	76.9%	75.7%	76.6%
***Ca.* Nitrotoga coloradensis SPKER**	87.7%	87.9%	87.5%	100.0%	73.9%	76.1%	78.2%	74.0%	75.0%
***Sideroxydans lithotrophicus* ES-1**	65.8%	66.1%	65.8%	64.5%	100.0%	76.4%	78.0%	78.4%	78.7%
***Gallionella capsiferriforma ns* ES-2**	63.2%	64.2%	63.7%	63.3%	66.1%	100.0%	76.7%	75.9%	78.2%
***Ca.* Gallionella acididurans ShG14-8**	64.6%	64.5%	64.6%	64.2%	67.1%	68.3%	100.0%	76.5%	77.1%
Ferriphaselus amnicola	65.1%	65.2%	65.1%	64.5%	67.5%	66.4%	65.3%	100.0%	83.7%
***Ferriphaselus* sp. R-1**	65.0%	65.3%	65.6%	64.6%	67.9%	66.0%	65.5%	84.7%	100.0%