Genomic profiling of four cultivated *Candidatus* Nitrotoga spp. predicts broad metabolic potential and environmental distribution

Andrew M. Boddicker and Annika C. Mosier

SUPPLEMENTAL 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 (based in part on Synthetic Crenarchaeota Media; Könneke et al., 2005). 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 (final pH ~7.1) 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.

Fluorescent *in situ* hybridization

Fluorescent in situ hybridization (FISH) (Daims *et al.,* 2005) was carried out on fixed (2% paraformaldehyde final concentration) cells filtered onto 0.2 µm Isopore polycarbonate membranes (Millipore, Burlington, MA). Filters were treated with 1M HCl for 1 hour at 37°C to enhance permeabilization. Oligonucleotide probes were EUB338 I-III (all Bacteria) and

Ntoga122 with competitor probes c1Ntoga122 and c2Ntoga122 (*Ca.* Nitrotoga) (Lücker et al., 2015). Filters were hybridized at 40% formamide for five hours and imaged on a Zeiss LSM 700 confocal microscope. Cell counts were converted to a percentage of the total cell count found using the general nucleic acid stain 4',6-diamidino-2-phenylindole (DAPI).

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 *Ca.* 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 *Ca.* Nitrotoga genus.

Putative *Ca.* Nitrotoga bins of interest were manually refined using the Anvi'o metagenomics pipeline version 2.1.0 (Eren *et al.*, 2015). Putative *Ca.* 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 *Ca.* 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 *Ca.* Nitrotoga genomes (from this study).

16S rRNA gene sequences (which rarely bin properly in metagenomic assemblies) were manually added to the *Ca.* 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 *Ca.* Nitrotoga hit of that size) to any 16S rRNA gene from the database were added to the *Ca.* 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 *Ca.* Nitrotoga genomes, only one 16S rRNA gene assembled on a contig with paired reads mapped to other *Ca.* Nitrotoga genomic contigs. The assembled 16S rRNA genes from each culture were most similar to *Ca.* Nitrotoga sequences in the SILVA database. The correct contig containing the *Ca.* 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 *Ca.* 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 *Ca.* 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 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 *Ca.* Nitrotoga *nxrA* genes had SNVs and the SPKER *nxrBCD* genes were intact with no SNVs. A consensus *nxrA* sequence is presented in this study for the SPKER genome, as paired reads could not resolve the SNVs into individual gene copies.

Relative abundance of *Ca.* 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 operational taxonomic units (OTUs) at 97% sequence identity and the DECIPHER web tool was used to check for chimeras (short sequences option) (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

Ca. Nitrotoga genome assembly

CheckM indicated that the *Ca.* 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 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.

Ca. Nitrotoga genomes were estimated to contain 0.24-0.3% contamination. Specifically, the CP45, LAW, MKT genomes had duplications of the same two marker genes (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.

***Ca.* 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)

Ca. Nitrotoga *nxr* genes were ultimately placed on single contigs within each genome forming an *nxrABCD* operon. *Ca.* 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). *Ca.* Nitrotoga NxrB (TIGR03478, evalue <2.6e-73) was 385 amino acids long in all genomes, while *Ca.* 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).

Genes encoding two conserved hypothetical proteins were found downstream of the *nxr* operon in most *Ca.* 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 gene for the 341 amino acid protein was found on a neighboring contig. The CP45 *nxr* contig had additional truncated genes for hypothetical proteins on either end of the contig.

Ca. Nitrotoga NXR subunits had a conserved protein structure that was highly divergent from other NXRs and members of the Type II DMSO reductase family of enzymes (Figure 4). *Ca.* Nitrotoga NxrA subunits were at least 98.1% identical to each other across the 1169 amino acid predicted 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 *Ca.* 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 gene encoding 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 *Ca.* Nitrotoga genomes.

Each *Ca.* Nitrotoga genome had a gene that encodes 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 *Ca.* 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 may 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, *Ca.* Nitrotoga may be capable of detoxifying cyanide while also contributing to their sulfur metabolism via sulfite production (see *sulfur energetics* section).

***Ca.* 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) genes, encoding the key enzyme for CO_2 fixation, while the SPKER genome had single copies. All *Ca.* Nitrotoga genomes had a predicted 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).

Genes encoding the two key enzymes of the reverse tricarboxylic acid (rTCA) cycle (oxoglutarate:ferredoxin oxidoreductase and ATP-citrate lyase) were missing from *Ca.* 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 *Ca.* 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 *Ca.* Nitrotoga genomes, except the LAW genome was missing a cellobiose phosphorylase gene.

Extracellular polymeric substance (EPS) synthesis is also likely in *Ca.* Nitrotoga, as an extensive array of PEP-CTERM exopolysaccharide sorting enzymes (Haft *et al.*, 2006),

polysaccharide export pumps, and saccharide modifier genes (i.e., UDP-glucose dehydrogenase, polyprenyl glycosylphosphotransferase, D-alanyl-lipoteichoic acid acyltransferase) were present in all genomes. EPS production has been observed in *Ca.* Nitrotoga via microscopy (Ishii *et al.*, 2017), and may explain the formation of microcolonies in WWTPs (Lücker *et al.*, 2015).

***Ca.* 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 *Ca.* 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 *Ca.* Nitrotoga genomes. However, genes for 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), were not detected in *Ca.* 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 zinc/cadmium transporter genes in *Ca.* 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 *Ca.* Nitrotoga genomes. The presence of two putative hemolysin genes in all *Ca.* Nitrotoga genomes, and an additional heme oxygenase in the CP45 genome, may support an antagonistic role of *Ca.* 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, Stoecker K, Wagner M. (2005). Fluorescence in situ hybridization for the detection of prokaryotes. In: Osborn AM, Smith CJ (eds) *Molecular Microbial Ecology*. Taylor & Francis: New York, NY, USA, pp 213–239.

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

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.

Könneke, M, Bernhard, A, de la Torre, J, Walker, C, Waterbury, J, Stahl, D. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437(7058):**543–546.

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.