Metagenomic profiling of ammonia- and methane-oxidizing microorganisms in a Dutch drinking water treatment plant

Lianna Poghosyan, Hanna Koch, Jeroen Frank, Maartje A.H.J. van Kessel, Geert Cremers, Theo van Alen, Mike S.M. Jetten, Huub J.M. Op den Camp, Sebastian Lücker*

Department of Microbiology, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

*Correspondence: s.luecker@science.ru.nl

Keywords

Sand filtration; nitrification; comammox *Nitrospira*; methanotrophic bacteria; metagenomics

Highlights

- Microbial distribution was mainly influenced by sampling location within the DWTP
- Clade A comammox *Nitrospira* were the dominant nitrifiers in the primary sand filter
- Clade B was most abundant in samples from wall biofilm and the secondary filter
- A novel *Methylophilaceae*-affiliated methanotroph dominated the primary sand filter

Abstract

Elevated concentrations of ammonium and methane in groundwater can cause severe problems during drinking water production. To avoid their accumulation, raw water in the Netherlands, and many other countries, is purified by sand filtration. These drinking water filtration systems select for microbial communities that mediate the biodegradation of organic and inorganic compounds. In this study, the active layers and wall biofilm of a Dutch drinking water treatment plant (DWTP) were sampled at different locations along the filtration units of the plant over three years. We used high-throughput sequencing in combination with differential coverage and sequence composition-based binning to recover 56 near-complete metagenome-assembled genomes (MAGs) with an estimated completion of ≥70% and with ≤10% redundancy. These MAGs were used to characterize the microbial communities involved in the conversion of ammonia and methane. The methanotrophic microbial communities colonizing the wall biofilm (WB) and the granular material of the primary rapid sand filter (P-RSF) were dominated by members of the *Methylococcaceae* and *Methylophilaceae.* The abundance of these bacteria drastically decreased in the secondary rapid sand filter (S-RSF) samples. In all samples, complete ammonia-oxidizing (comammox) *Nitrospira* were the most abundant nitrifying guild. Clade A comammox *Nitrospira* dominated the P-RSF, while clade B was most abundant in WB and S-RSF, where ammonium concentrations were much lower. In conclusion, the knowledge obtained in this study contributes to understanding the role of microorganisms in the removal of carbon and nitrogen compounds during drinking water production. We furthermore found that drinking water treatment plants represent valuable model systems to study microbial community function and interaction.

40 **Introduction**

41 About 97% of all available water on earth is saline. The remaining 3% is freshwater, of which
42 more than two-thirds is frozen in ice sheets. Thus, only a small fraction of the global freshwater
43 exists as ground and surface water that is available for drinking water production. According
44 to the European Commission (EC, 2016), about 50% of drinking water in Europe is produced
45 from groundwater and 37% from surface water.

46 Groundwater has a relatively constant composition and may contain high concentrations of
47 iron (Fe^{2+}; 0.9-7.8 mg/L), manganese (Mn^{2+}; 0-0.56 mg/L), ammonium (NH_4^+; 0.1-0.5 mg/L),
48 and some organic compounds such as methane (CH_4; 0-37 mg/L) (Albers et al., 2015; Li and
49 Carlson, 2014; Osborn et al., 2011). Elevated concentrations of these compounds in
50 groundwater can cause severe problems during drinking water production and distribution
51 (Okoniewska et al., 2007; Rittmann et al., 2012; Sharma et al., 2005). Biofiltration (e.g., rapid
52 (RSF) or slow (SSF) sand filtration, granular activated carbon filters) are widely applied
53 methods for the removal of the above-mentioned compounds. Biofilters harbor complex
54 microbial communities that are introduced via the source water (Yang et al., 2016) and are
55 shaped by the configuration of the treatment process (Li et al., 2017; Pinto et al., 2012). In the
56 filtration units microbial growth is stimulated on filter material, mediating the biodegradation
57 of organic and inorganic compounds (Proctor and Hammes, 2015). Gases such as methane,
58 hydrogen sulfide, carbon dioxide, and other volatile compounds are removed from the
59 groundwater through gas exchange systems (Trussell et al., 2012). The increased dissolved
60 oxygen in the water caused by this mechanical aeration step serves as an electron acceptor in
61 microbially mediated oxidative reactions, which may ensure the near-complete nutrient
62 removal in the biologically active layer of the sand filters.

63 One of the main groundwater contaminants is ammonium. Excess ammonium in raw water is
64 often associated with microbiological, chemical and sanitary problems in drinking water

65 distribution systems, such as excessive biofilm growth, pH decrease, pipe corrosion, and
66 elevated nitrite and nitrate levels (Beech and Sunner, 2004; Camper, 2004; Rittmann et al.,
67 2012). In engineered systems, such as drinking water treatment plants (DWTP), ammonium
68 removal is achieved by the activity of nitrifying microorganisms that oxidize ammonia to
69 nitrate via a series of intermediates. While canonical nitrifying guilds perform ammonia- and
70 nitrite-oxidation in a tight interplay, complete ammonia-oxidizing (comammox) bacteria of the
71 genus *Nitrospira* possess all proteins necessary to perform nitrification on their own (Daims et
72 al., 2015; van Kessel et al., 2015). Nitrifying microbial communities of rapid sand filters have
73 been studied before and seem to be represented by different groups of nitrifiers (Albers et al.,
74 2015; Fowler et al., 2018; Gülay et al., 2016; Oh et al., 2018; Palomo et al., 2016; Pinto et al.,
75 2015; van der Wielen et al., 2009). In these systems, nitrification can be limited by the
76 availability of vital nutrients for this process, such as phosphate and copper. This can cause
77 incomplete nitrification (de Vet et al., 2012; Wagner et al., 2016), leading to incomplete
78 ammonium removal and/or nitrite accumulation (Wilczak et al., 1996), and microbial after-
79 growth in the distribution network (Rittmann et al., 2012).
80 Methane is a colorless and odorless gas that usually does not present a health risk in drinking
81 water. However, methane gas is highly flammable and can be explosive at elevated
82 concentrations, and it also can serve as substrate for growth of microorganisms in distribution
83 systems. Most of the methane is removed by the mechanical aeration step and remaining
84 amounts are oxidized by aerobic methanotrophs within the sand beds of the filter units.
85 Correspondingly, a number of studies reported methane-oxidizing bacteria colonizing the
86 granular material of the sand filters (Albers et al., 2015; Gülay et al., 2016; Palomo et al., 2016).
87 However, in contrast to nitrifying microbial communities, methanotrophs in these engineered
88 systems are traditionally less well studied. Methane is a potent greenhouse gas, with a global
89 warming potential of 34 CO_2 equivalents over 100 years (Stocker et al., 2013) and the removal

90 of most methane from the drinking water via aeration causes methane emissions to the
91 atmosphere, contributing to global warming and climate change (Maksimavičius and Roslev,
92 2020).
93 Like in many European countries, groundwater is the primary source (65%) for drinking water
94 production in the Netherlands (Dutch Drinking Water Statistics 2017). In this study, we used
95 genome-resolved metagenomics and gene-centric approaches to analyze microbial
96 communities in a Dutch DWTP, with a special focus on ammonia- and methane-oxidizing
97 microorganisms. The groundwater entering this DWTP contains elevated concentrations of
98 organic (CH_4, 5.2 mg/L) and inorganic (NH_4^+, 0.66 mg/L; Fe^{2+}, 8.4 mg/L; Mn^{2+}, 0.18 mg/L)
99 nutrients, which are removed in two sequential rapid sand filters. Thus, this DWTP represents
100 an interesting model system to study microbial communities involved in the conversion and
101 removal of these compounds. Metagenomic analysis of samples taken at different stages within
102 this DWTP revealed the key microorganisms involved in ammonia- and methane-oxidation,
103 including novel methanotrophic bacteria from the *Methylophilaceae* family, which were
104 assumed previously to comprise only methylotrophic bacteria.

105 **Materials and Methods**

106 *Sample collection and trace element analysis*

107 Samples were obtained from the pumping station Breehei, a drinking water treatment facility
108 located in Venray, the Netherlands (51°28'54.6"N; 5°59'10.2"E), operated by NV Waterleiding
109 Maatschappij Limburg. Drinking water is produced from groundwater. Samples from the
110 active layers of primary and secondary rapid sand filters (P-RSF and S-RSF) were collected in
111 50 ml sterile falcon tubes in June 2016 and September 2018. Samples from the biofilm formed
112 on the walls of the primary sand filter (WB) were taken in June 2016, May 2017 and September
113 2018 (Figure 1). All samples were transferred to the laboratory within 4 h, and were stored at

4°C for further analysis. Water quality parameters were determined by Aqualab Zuid (https://www.aqualab.nl) over the period 2000-2016 (Table 1).

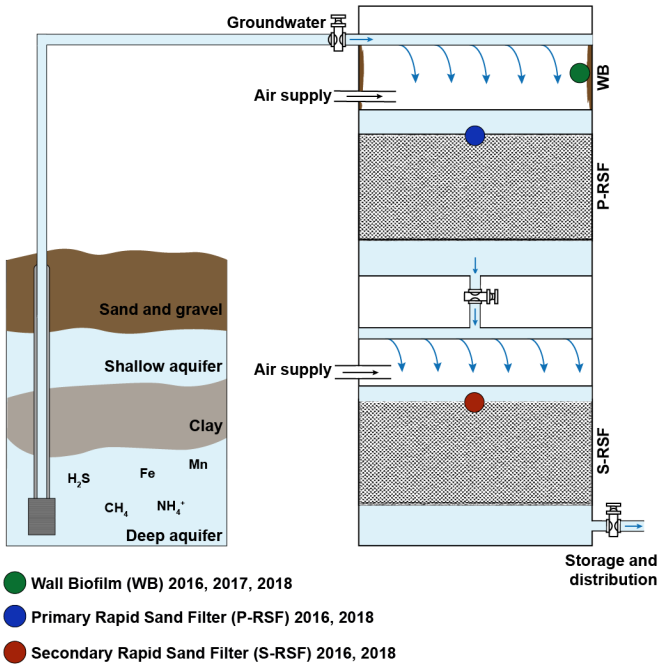


Figure 1. Schematic illustration of DWTP Breehei. Sampling locations are indicated by colored dots, sampling timepoints (in years only) for each location are indicated.

Table 1. Average water quality parameters of the incoming groundwater and at different stages along the treatment train.

Unit (mg L^{-1})	Ground-water	After aeration	P-RSF effluent	S-RSF effluent	Drinking water	WML standard[a]
Methane (CH_4)	5.2	1.2	0.012	0.012	0.012	0.5
Ammonium (NH_4^+)	0.66	0.65	0.0145	0.039	0.039	0.1
Nitrite (NO_2^-)	0.0046	*	0.063	0.0082	0.0041	0.05
Nitrate (NO_3^-)	0.13	*	*	*	1.64	40
Oxygen (O_2)	0.4	9.1	5.0	10.2	10.2	-
Carbon dioxide (CO_2)	42.38	*	18	6.16	6.95	-
Iron (Fe)	8.4	*	0.11	0.053	0.043	0.1
Manganese (Mn)	0.18	*	0.022	0.0079	0.0052	0.025

* Not applicable or were not analyzed

\- No standards

[a]Defined by Waterleiding Maatschappij Limburg

Methane uptake

Methane-oxidizing capacity was determined for the P-RSF and WB samples collected in 2016. For the P-RSF, 2.5, 5, 10, and 20 g of sand material were mixed with 20 ml top water. After

127 settling of the sand, overlaying water samples (20 ml) were transferred into 120 ml serum
128 bottles with ±14.5 mg/L CH_4 in the headspace and incubated at room temperature in a shaking
129 incubator (200 rpm). For the WB sample, the incubations were conducted in triplicates using
130 1 and 2.5 g of biomass. CH_4 concentrations were determined by sampling 300 µL headspace,
131 which were injected in triplicates into a HP 5890 gas chromatograph (Hewlett Packard, Palo
132 Alto, CA).

133 *DNA extractions*

134 DNA from samples for Illumina sequencing was extracted using two different methods to
135 obtain differential abundance information. DNA from 2016 samples was extracted with
136 DNeasy Blood & Tissue Kit (Qiagen Ltd., West Sussex, United Kingdom) and PowerSoil DNA
137 Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA). Samples (0.5 g) were mechanically
138 disrupted using a TissueLyser (Qiagen) for 2 x 30 seconds at 30 Hz, followed by DNA
139 extractions according to the manufacturer's instructions. For the extraction of DNA from
140 samples collected in 2017 and 2018, the DNeasy Blood & Tissue kit was replaced by
141 ammonium acetate (Kowalchuk et al., 2004) and CTAB (Zhou et al., 1996) based extraction
142 methods, respectively.
143 For long read Nanopore sequencing, DNA was extracted from P-RSF samples collected in
144 2016 and 2018 using the CTAB-based extraction method. To avoid shearing of genomic DNA,
145 all bead-beating and vortexing steps were replaced by carefully inverting the tubes several
146 times, and all the pipetting steps were performed using cut-off pipette tips. Genomic DNA was
147 purified twice by phenol:chloroform:isoamyl alcohol (25:24:1) phase extraction (Zhou et al.,
148 1996). Extracted DNA was resuspended in nuclease-free water and stored at 4°C.

149 *Illumina library preparation and sequencing*

150 For Illumina library preparation, the Nextera XT kit (Illumina, San Diego, CA, USA) was used
151 according to the manufacturer's instructions. Enzymatic tagmentation was performed using 1
152 ng of DNA per sample, followed by incorporation of the indexed adapters and library
153 amplification. After subsequent purification using AMPure XP beads (Beckman Coulter,
154 Indianapolis, IN, USA), libraries were checked for quality and size distribution using the 2100
155 Bioanalyzer with the High Sensitivity DNA kit (Agilent, Santa Clara, CA, USA). Library
156 quantitation was performed by Qubit using the Qubit dsDNA HS Assay Kit (Thermo Fisher
157 Scientific, Waltham, MA, USA). After dilution to 4 nM final concentration, the libraries were
158 pooled, denatured, and sequenced on an Illumina Miseq. Paired-end sequencing of 2 x 300
159 base pairs was performed using the Illumina MiSeq Reagent Kit v3 according to the
160 manufacturer's protocol.

161 *Nanopore library preparation and sequencing*

162 For Nanopore library preparation, 1 - 1.5 µg of DNA, measured by Qubit with the Qubit
163 dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA), was used. The input
164 DNA was quality-checked by agarose gel electrophoresis to contain only high molecular DNA
165 and show no degradation. For sequencing, DNA Library construction was performed using the
166 Ligation Sequencing Kit 1D (SQK-LSK108) in combination with the Native Barcoding
167 Expansion Kit (EXP-NBD103 or EXP-NBD104) according to the manufacturer's protocol
168 (Oxford Nanopore Technologies, Oxford, UK). Fragments were end-repaired using the
169 NEBNext® FFPE DNA Repair Mix (New England Biolabs, Ipswich, MA, USA), with
170 subsequent fragment purification using AMPure XP beads (Beckman Coulter Life Sciences,
171 Indianapolis, IN, USA). End repair and dA-tailing was done using the NEBNext® Ultra™ II
172 End Repair/dA-Tailing Module (New England Biolabs) followed by a cleanup of the fragments

173 using AMPure XP beads. Selected barcodes for each sample were ligated using the Blunt/TA
174 Ligase Master Mix (New England Biolabs) and the resulting fragments were purified using
175 AMPure XP beads. The DNA concentration of all libraries was measured by Qubit using the
176 dsDNA HS Assay Kit and pooled to a maximum of 700 ng DNA. Subsequently, adapters were
177 ligated using the NEBNext® Quick Ligation Module (New England Biolabs). After purification
178 using AMPure XP beads, the pooled libraries were quantified again using Qubit. The libraries
179 were loaded on a Flow Cell (R9.4.1) and run on a MinION device (Oxford Nanopore
180 Technologies, Oxford, UK), according to the manufacturer's instructions. Base calling after
181 sequencing was done using Albacore v2.1.10, (Oxford Nanopore Technologies) for the 2016
182 and guppy_basecaller in combination with guppy_barcoder (Oxford Nanopore Technologies,
183 Limited Version 2.3.7+e041753) for the 2018 sample.

184 *Assembly*

185 Raw Illumina sequence data were processed for quality-trimming, adapter removal, and
186 contamination-filtering, using BBDUK (BBTOOLS v37.17; http://jgi.doe.gov/data-and-
187 tools/bbtools/bb-tools-user-guide/bbduk-guide/). All sequencing reads from the same sampling
188 location were co-assembled using metaSPAdes v3.10.1 (Nurk et al., 2017) with the following
189 parameters: k-mer sizes 21, 33, 55, 77, 99 and 127, minimum contig length 1500 bp. Raw
190 Nanopore reads were quality trimmed using Filtlong v0.2.0
191 (https://github.com/rrwick/Filtlong) with minimum read length 1000 bp and error rate <20%.
192 Porechop v0.2.4 (https://github.com/rrwick/Porechop) was used to remove adapters and split
193 chimeric reads with default settings. Trimmed reads were assembled using Canu v1.8 (Koren
194 et al., 2017) with minimum read length 1000, corrected read error rate 0.105, and genome size
195 5m. To correct error rates, the trimmed Nanopore reads were mapped to the assembly using
196 Minimap2 v2.16-r922 (Li, 2018), which then was polished with Racon v1.3.1 (Vaser et al.,

197 2017). Subsequently, Racon v1.3.1 was used to further polish the assembly twice with Illumina
198 reads obtained from the 2018 P-RSF sample.

199 *Metagenome binning*

200 Differential coverage information was determined by separately mapping the sequencing reads
201 from each sample and DNA extraction method against the obtained co-assemblies, using
202 Burrows-Wheeler Aligner v0.7.15 (BWA) (Li and Durbin, 2010) and employing the "mem"
203 algorithm. For Canu assembly, only the Illumina reads from P-RSF 2018 samples were
204 mapped. The generated sequence alignment map (SAM) files were converted to binary format
205 (BAM) using SAMtools (Li et al., 2009). Metagenome binning was performed using anvi′o
206 v5.3 (Eren et al., 2015). Anvi′o′s Snakemake-based (Koster and Rahmann, 2012) contigs
207 workflow was used to generate contig databases from each Illumina and Nanopore
208 assembly. Shortly, anvi′o employs Prodigal v2.6.3 (Hyatt, 2010) to identify open reading
209 frames (ORFs) and HMMER v3.2 (Eddy, 2011) to identify archaeal (Rinke et al., 2013) and
210 bacterial (Campbell et al., 2013) single-copy core genes. The Cluster of Orthologous Groups
211 of proteins (COG) database (Tatusov et al., 1997) together with Centrifuge v1.0.3-beta (Kim
212 et al., 2016) was used to annotate genes in the contig databases. Each BAM file was profiled
213 to obtain differential coverage and statistical information based on mapping results and to
214 generate merged profile databases. The contigs were then automatically clustered with
215 CONCOCT (Alneberg et al., 2014), followed by manual binning and bin refinement using
216 the anvi′o interactive interface (Delmont and Eren, 2016; Eren et al., 2015). Completeness
217 and contamination (referred to as redundancy in this study) of bins was assessed by CheckM
218 v1.01.11 (Parks et al., 2015), which uses pplacer v1.1 alpha 19 (Matsen et al., 2010) to
219 identify and quantify single-copy marker genes. Based on the suggested standards (Bowers
220 et al., 2017), the bins were defined as high-quality (>90% complete and <5% redundancy,

221 complete small subunit rRNA operon, ≥18 tRNAs) and medium-quality (≥70% complete
222 and <10% redundancy) metagenome-assembled genomes (MAGs).

223 *Dereplication and taxonomic classification of MAGs*

224 MAGs were dereplicated using dRep v2.2.3 (Olm et al., 2017) at 99% average nucleotide
225 identity (ANI) for clustering. Within each cluster, the best MAG was selected based on
226 completeness (≥70%), redundancy (<10%), N50 of contigs, and fragmentation. GTDB-Tk
227 v0.3.2 (Parks et al., 2018) was used for taxonomic assignment of the final MAGs. Phyla are
228 named according to the recently suggested nomenclature (Whitman et al., 2018) using
229 standardized phylum suffix -ota.

230 *Abundance estimation of MAGs*

231 To calculate the relative abundance of the dereplicated MAGs in each sample, reads from
232 all samples were individually mapped to each co-assembly using BWA v0.7.15 (Li and
233 Durbin, 2010) as described above. The coverage of each MAG was calculated using CheckM
234 v1.01.11 (minimum alignment length 0.95) (Parks et al., 2015) and was normalized by
235 multiplying this coverage with a normalization factor (sequencing depth of the largest
236 sample divided by the sequencing depth of each individual sample). The distribution of
237 MAGs was calculated as percentages by dividing a MAG's coverage in each sample by the
238 total coverage of the respective MAG in all samples.

239 *Functional analysis*

240 For the gene-centric approach, co-assembly across all samples was performed using
241 MEGAHIT v1.1.1-2 (Li et al., 2015). Open reading frames (ORFs) in the MAGs obtained
242 above and the new co-assembly were predicted using Prodigal v2.6.3 (Hyatt, 2010), which was
243 set to include partial ORFs. Custom-build hidden Markov models (HMMs) (Eddy, 2011) of

244 specific marker proteins were used (Supplementary material and methods) to annotate all ORFs
245 using hmmsearch (HMMER v3.1b2; http://hmmer.org). The HMM for RNA polymerase
246 subunit beta (RpoB) was downloaded from FunGene (Fish et al., 2013). Remaining ORFs in
247 the MAGs were annotated using Prokka v1.12-beta (Seemann, 2014). The annotations of all
248 genes discussed in this study were confirmed by BLAST against the TrEMBL, Swiss-Prot and
249 NCBI nr databases. Subcellular localization of the proteins was predicted by SignalP 5.0
250 (Armenteros et al., 2019) and TMHMM 2.0 (Krogh et al., 2001).
251 Functional gene-based abundances of ammonia- and methane-oxidizing microorganisms were
252 estimated using competitive metagenomic read recruitment to ensure unique mapping. For this,
253 reads from each metagenomic sample were mapped using bowtie2 v2.3.1 (Langmead and
254 Salzberg, 2012) in '-very-sensitive' mode against extracted partial and complete sequences of
255 *rpoB* and the ammonia as well as the particulate and soluble methane monooxygenase subunit
256 A/alpha genes (*amoA*, *pmoA* and *mmoX*, respectively). SAMtools flagstat v1.6 (Li et al., 2009)
257 was used to obtain the number of mapped reads. Reads per kilo base per million mapped reads
258 (RPKM)-values were used to correct for differences in sequencing depth and gene length. To
259 estimate the relative abundance of microorganisms encoding ammonia and methane
260 monooxygenases in each sample, the normalized read counts were calculated as fraction of
261 the normalized read counts of the identified *rpoB* genes.

262 *Phylogenomic and phylogenetic analyses*

263 The up-to-date bacterial core gene (UBCG) pipeline (Na et al., 2018) with default parameters
264 was used to extract and concatenate bacterial core gene sets. To infer the phylogeny of the
265 archaeal MAGs, the anvi'o phylogenomic workflow
266 (http://merenlab.org/2017/06/07/phylogenomics) was used to individually align and
267 concatenate 162 archaeal single-copy genes. Maximum-likelihood trees were calculated using
268 RAxML version 8.2.10 (Stamatakis, 2014) on the CIPRES science gateway (Miller et al.,

269 2010). For details, see Supplementary Methods. Amino acid sequences of type II DMSO
270 reductase-family enzymes, and of ammonia and methane monooxygenases were aligned using
271 ARB v5.5 (Ludwig et al., 2004). The maximum-likelihood trees were calculated using RAxML
272 HPC-HYBRID v.8.2.12 on the CIPRES and IQ-tree webserver (Trifinopoulos et al., 2016) as
273 described in Supplementary Methods. All phylogenetic trees were visualized in iTOL (Letunic
274 and Bork, 2016).

275 *Data visualization*

276 Manuscript figures were generated using ggplot2 (Wickham, 2016), Rstudio (Racine, 2012)
277 and the anvi′o interactive interface (http://merenlab.org/2016/02/27/the-anvio-interactive-
278 interface).

279 **Results**

280 *DWTP performance*

281 The produced water quality analyses (Table 1) showed that methane, ammonium, nitrite and
282 nitrate were well below the required quality standards. Most of the methane in the raw water
283 was removed during the aeration step and the remainder was oxidized in the P-RSF.
284 Pronounced ammonia oxidation in P-RSF resulted in nitrite accumulation in the effluent of this
285 filter, which was subsequently removed in the S-RSF. Similarly, the Fe^{2+} and Mn^{2+} removal
286 efficiency of the system was very high (>99%). In addition, we determined methane uptake
287 rates in the P-RSF and WB samples. Complete methane oxidation was achieved in all samples
288 within 5 to 8 days of incubation, except for the control containing only raw water (Figure S1).
289 Methane consumption in P-RSF samples increased with an increasing amount of sand material,
290 whereas no significant difference in methane uptake was detected between the incubations with
291 1 and 2.5 g of WB biomass (Figure S1). The identical oxidation rates in the incubations with

292 WB biomass seem counterintuitive, but may be caused by sample inhomogeneity and uneven
293 distribution of methanotrophic bacteria within the biofilm.

294 *Recovery of metagenome-assembled genomes*

295 Over the period 2016 to 2018 a total of 7 samples from DWTP Breehei were collected (Figure
296 1) and sequenced. This resulted in a total of 125 million paired-end Illumina sequencing reads.
297 For each sample location more than 70% of the respective reads could be co-assembled into
298 ~413, 184, and 249 Mb sequencing data for WB, P-RSF, and S-RSF, respectively (Table S1).
299 In addition, P-RSF samples were also sequenced using the Oxford Nanopore long-read
300 platform to improve the assembly of the most abundant microorganisms. Overall, binning of 4
301 individual metagenome assemblies based on sequence composition and differential coverage
302 patterns resulted in 78 near-complete Illumina and 7 Nanopore MAGs (Table S2). All MAGs
303 obtained from Illumina co-assemblies as well as the Nanopore assembly were dereplicated at
304 strain level (99% ANI), which yielded 50 medium and 6 high-quality (MIMAG standards;
305 Bowers et al., 2017) non-redundant MAGs (Table 2, Table S2) that were used for downstream
306 analyses. Given the high number of single nucleotide variants observed during bin refinement
307 using the anvi′o interactive interface, 15 MAGs were categorized as population-level genomes
308 (Table 2). All MAGs were classified at the lowest possible taxonomic level using GTDB-tk
309 (Parks et al., 2018), indicating affiliation with 1 archaeal and 12 different bacterial phyla
310 (Figure 2, Table 2).

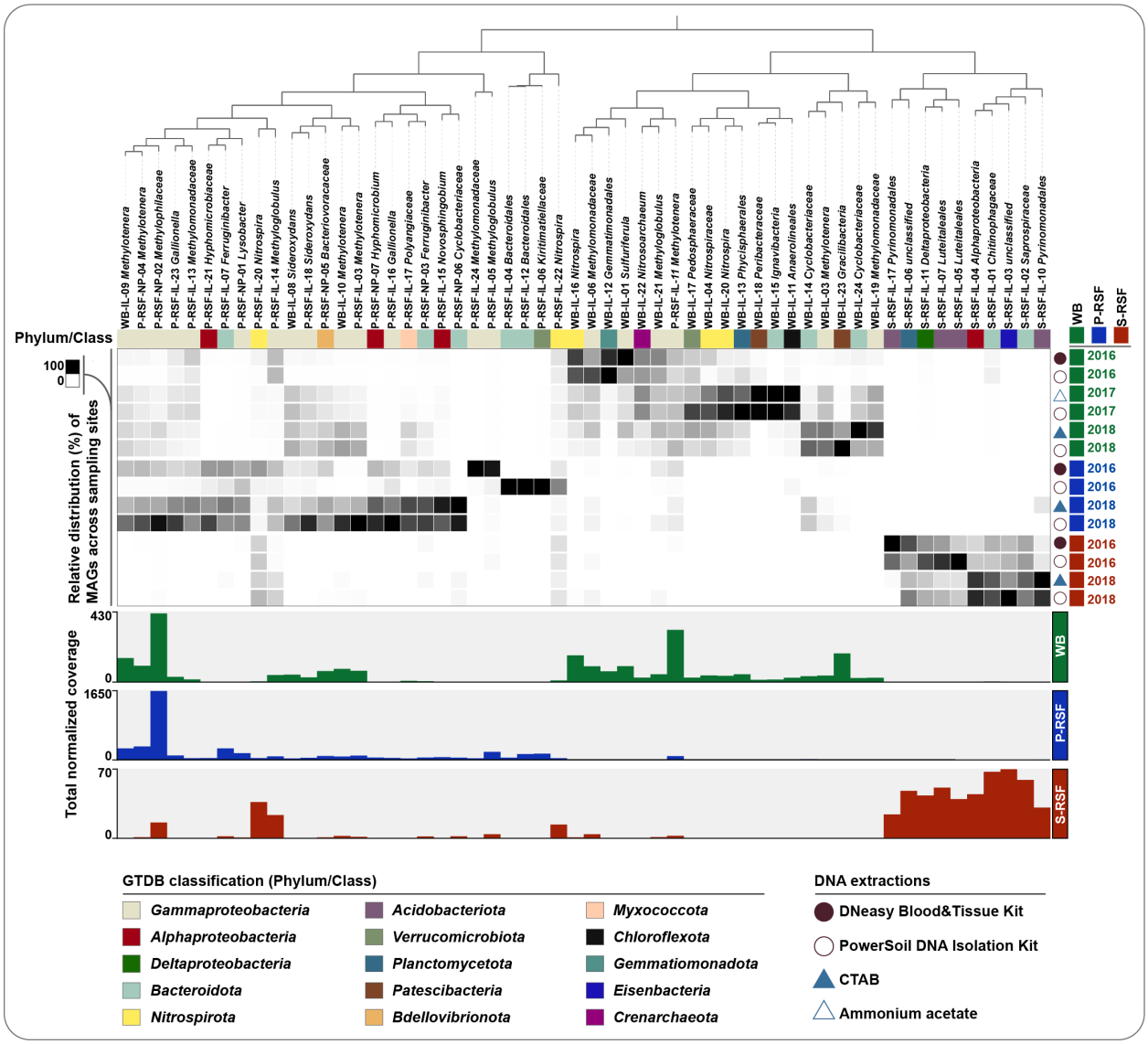


Figure 2. Overview of the distribution, coverage and classification of 56 dereplicated medium- and high-quality MAGs recovered from the DWTP Breehei metagenomes. MAGs are organized based on their relative distribution in the DWTP sample locations using Euclidean distance and Ward linkage as implemented in the anvi′o interactive interface. The phylogenetic affiliation of the MAGs based on GTDB classification is indicated by colored boxes. The heatmap indicates the percent coverage of each MAG in a given metagenome. The bar plots show the total normalized coverage of each MAG per sampling site.

Table 2. General characteristics of MAGs recovered from DWTP Breehei metagenomes.

MAG ID[a]	Classification[b]	Complete-ness (%)[c]	Redun-dancy (%)[c]	Length (Mb)	N50	Number contigs	MIMAG[d] quality
P-RSF-IL-03	g_*Methylotenera*	99.57	0.43	2.48	204,148	20	Medium
P-RSF-IL-04	o_*Bacteroidales*	95.70	1.09	4.38	46,595	154	Medium
P-RSF-IL-05	g_*Methyloglobulus*	99.25	0.82	3.87	67,388	99	Medium
P-RSF-IL-06	f_*Kiritimatiellaceae*	93.41	1.38	2.21	572,251	7	High
P-RSF-IL-07	g_*Ferruginibacter*	96.50	1.72	3.05	104,094	43	Medium

P-RSF-IL-11[a]	g_*Methylotenera*	98.18	0.96	2.48	30,410	142	Medium
P-RSF-IL-12	o_*Bacteroidales*	98.57	3.28	5.54	156,716	53	Medium
P-RSF-IL-13	f_*Methylomonadaceae*	95.12	1.33	3.91	14,718	373	Medium
P-RSF-IL-14[a]	g_*Methyloglobulus*	76.88	6.08	3.76	6,341	643	Medium
P-RSF-IL-15[a]	g_*Novosphingobium*	90.61	0.92	2.43	12,194	277	Medium
P-RSF-IL-16	g_*Gallionella*	85.00	2.45	2.00	12,610	203	Medium
P-RSF-IL-17	f_*Polyangiaceae*	85.52	6.71	9.19	8,612	1283	Medium
P-RSF-IL-18	g_*Sideroxydans*	97.36	5.02	2.47	19,722	202	Medium
P-RSF-IL-20	g_*Nitrospira*	90.17	3.45	3.85	9,698	498	Medium
P-RSF-IL-21	f_*Hyphomicrobiaceae*	83.07	1.45	3.92	8,625	562	Medium
P-RSF-IL-22[a]	g_*Nitrospira*	74.42	1.95	2.60	5,252	519	Medium
P-RSF-IL-23[a]	g_*Gallionella*	73.19	5.24	1.91	6,356	338	Medium
P-RSF-IL-24	f_*Methylomonadaceae*	91.70	2.96	3.55	13,253	354	Medium
P-RSF-NP-01	g_*Lysobacter*	99.95	1.45	4.15	4,155,120	1	High
P-RSF-NP-02[a]	f_*Methylophilaceae*	97.86	1.71	3.20	2,013,186	4	High
P-RSF-NP-03	g_*Ferruginibacter*	94.09	0.00	5.44	2,880,874	5	Medium
P-RSF-NP-04[a]	g_*Methylotenera*	91.45	0.88	2.48	542,553	11	High
P-RSF-NP-05	f_*Bacteriovoracaceae*	90.62	1.84	3.58	236,271	21	High
P-RSF-NP-06	f_*Cyclobacteriaceae*	98.36	0.55	4.64	3,263,186	2	Medium
P-RSF-NP-07[a]	g_*Hyphomicrobium*	94.04	2.42	3.94	80,300	69	Medium
S-RSF-IL-01	f_*Chitinophagaceae*	98.03	0.25	3.85	72,704	83	Medium
S-RSF-IL-02	f_*Saprospiraceae*	98.77	0.64	3.97	34,568	196	Medium
S-RSF-IL-03	p_*Eisenbacteria*	96.15	1.10	3.80	25,445	215	High
S-RSF-IL-04	c_*Alphaproteobacteria*	94.77	2.33	4.03	18,338	319	Medium
S-RSF-IL-05	o_*Luteitaleales*	75.98	0.85	3.27	6,665	554	Medium
S-RSF-IL-06	p_*Planctomycetota*	81.72	4.15	4.12	6,813	693	Medium
S-RSF-IL-07	o_*Luteitaleales*	94.25	4.27	4.86	35,344	183	Medium
S-RSF-IL-10	o_*Pyrinomonadales*	74.62	1.28	3.49	7,085	579	Medium
S-RSF-IL-11	c_*Deltaproteobacteria*	84.08	3.36	5.77	9,024	795	Medium
S-RSF-IL-17	o_*Pyrinomonadales*	70.36	3.42	6.08	5,250	1235	Medium
WB-IL-01	g_*Sulfuriferula*	97.63	0.24	2.17	89,907	40	Medium
WB-IL-03	g_*Methylotenera*	98.72	2.14	2.61	26,003	159	Medium
WB-IL-04	f_*Nitrospiraceae*	96.76	3.18	3.58	54,739	105	Medium
WB-IL-06	f_*Methylomonadaceae*	99.88	0.45	4.27	93,669	82	Medium
WB-IL-08	g_*Sideroxydans*	98.57	2.30	2.50	25,161	145	Medium
WB-IL-09[a]	g_*Methylotenera*	88.16	1.96	2.16	16,133	197	Medium
WB-IL-10[a]	g_*Methylotenera*	90.61	1.71	2.11	17,763	172	Medium
WB-IL-11[a]	o_*Anaerolineales*	90.91	0.91	3.93	42,863	132	Medium
WB-IL-12	o_*Gemmatimonadales*	93.41	2.75	3.81	15,101	349	Medium
WB-IL-13	o_*Phycisphaerales*	94.89	2.27	5.04	34,488	268	Medium
WB-IL-14	f_*Cyclobacteriaceae*	93.03	1.74	4.77	12,449	498	Medium
WB-IL-15	c_*Ignavibacteria*	89.05	1.38	3.73	10,944	428	Medium
WB-IL-16[a]	g_*Nitrospira*	88.52	5.00	3.54	11,765	422	Medium
WB-IL-17	f_*Pedosphaeraceae*	95.83	6.76	5.57	27,453	284	Medium
WB-IL-18	f_*Peribacteraceae*	77.59	2.42	1.51	15,391	130	Medium
WB-IL-19	f_*Methylomonadaceae*	85.23	4.06	3.60	8,377	507	Medium
WB-IL-20[a]	g_*Nitrospira*	92.44	2.32	3.78	41,484	143	Medium
WB-IL-21	g_*Methyloglobulus*	83.76	5.07	3.38	6,990	545	Medium
WB-IL-22[a]	g_*Nitrosoarchaeum*	84.63	3.64	1.21	6,275	213	Medium
WB-IL-23[a]	c_*Gracilibacteria*	85.39	1.12	1.36	116,994	25	Medium
WB-IL-24	f_*Cyclobacteriaceae*	76.45	2.08	3.03	6,398	532	Medium

* ID indicates from which co-assembly (assembly) the MAGs were recovered. IL – Illumina co-assemblies, NP – Nanopore assembly

[a]Population level genome

[b]Based on GTDB classification (Parks et al., 2018)

[c]Based on lineage-specific marker sets determined with CheckM (Parks et al., 2015)

[d]Defined by (Bowers et al., 2017)

326 *Distribution and taxonomic composition of the DWTP microbiome*

327 The influence of sampling location, sampling time, and DNA extraction method on the
328 distribution of recovered MAGs was analyzed by hierarchical clustering using Euclidean
329 distance metrics with a Ward linkage algorithm, allowing grouping of the MAGs based on their
330 occurrence patterns in the different samples (Figure 2). Overall, the choice of a DNA extraction
331 method had no pronounced influence on the distribution of MAGs across samples, except for
332 the samples from P-RSF in 2016. In this specific sample, the normalized coverage values
333 indicated a strong extraction bias for DNA extracted using the Power soil (P-RSF16_PS)
334 compared to the Blood and Tissue kit (P-RSF16_BT; Table S2). Notably, this bias was not
335 observed for the 2016 S-RSF samples extracted with the same kits, or for any other DNA
336 extraction performed using the Power soil kit. It thus is difficult to conceive that the extraction
337 method was affecting specific members only in the 2016 P-RSF microbial community, and the
338 reason for the observed bias remains unclear. Overall, the sampling location had the most
339 substantial effect on microbial diversity and abundance. While some MAGs recovered from
340 WB and P-RSF were also present in S-RSF, the overall community of the S-RSF community
341 clearly differed from the other sampling locations (Figure 2). The microbial communities of
342 WB and P-RSF were generally more similar, but sampling location and time influenced the
343 relative abundance of MAGs across all samples.

344 To gain insights into the overall microbial community structure and diversity of the DWTP
345 Breehei, 16S rRNA gene sequences were retrieved directly from the metagenomic assemblies.
346 Both full-length and partial 16S rRNA gene sequences were used for further analyses, since
347 only 21 to 32% of 16S rRNA reads were assembled into full-length sequences (Figure S2).
348 Subsequently, microbial community composition was analyzed at the phylum and family levels
349 (Figure S3). Although changes in abundance were observed, phylum level classification
350 revealed no differences in microbial community composition between the sampling locations

351 (Figure S3-A). At the family level, the S-RSF microbial community was clearly distinct from
352 the WB and P-RSF community (Figure S3-B), corroborating the MAG co-occurrence pattern-
353 based observations when samples were organized using Euclidian distance and Ward
354 ordination (Figure 2).
355 16S rRNA gene-based taxonomic profiling identified 14 phyla with relative abundances ≥1%
356 (Figure S3A). *Gammaproteobacteria* (17-67%), *Bacteroidota* (5-17%), *Acidobacteria* (2-
357 15%), *Alphaproteobacteria* (4-12%), *Planctomycetota* (2-10%) *"Ca.* Patescibacteria" (CPR;
358 1-7%), and *Nitrospirota* (2-6%) were the most dominant phyla identified in all samples (Figure
359 S3-A). In most cases, this 16S rRNA gene-based analysis corresponded well to the taxonomic
360 affiliation of the recovered MAGs. Additionally, one MAG belonging to the candidatus phylum
361 "*Ca.* Eisenbacteria" and one *Crenarchaeota* (*Thaumarchaeota*) MAG were obtained from WB
362 and S-RSF (Figure 1), but were of low abundance in the 16S rRNA datasets.

363 *Genome functional profiling*

364 Iron, manganese, reduced sulfur species, ammonium and methane are removed during the
365 drinking water treatment process. However, our understanding of the microbial and
366 geochemical processes contributing to removal of these compounds during sand filtration is
367 still limited. Although iron (Fe^{2+}) oxidation pathways in bacteria are not fully understood, it is
368 known that certain autotrophic bacteria are responsible for this process. Four *Gallionellaceae*
369 MAGs (2 *Gallionella,* 2 *Sideroxydans*) were present only in WB and P-RSF metagenomes
370 (Figure 2), indicating that complete iron removal may already occur in P-RSF. The potential
371 to use reduced sulfur compounds as electron donor was encoded in *Hyphomicrobiaceae* (P-
372 RSF-IL-21) and *Sulfuriferula* (WB-IL-01) MAGs (Table S3). To learn more about the
373 microorganisms driving the removal of methane and ammonium in this rapid sand filtration
374 system, a genome-resolved metagenomics approach was applied to recover high-quality
375 genomes of the key ammonia- and methane-oxidizing microorganisms. Subsequently, all

recovered MAGs were screened for functional marker proteins involved in nitrification, and methane and other one-carbon (C1) compound oxidation (Figure 3, Table S3) using custom-build HMM models.

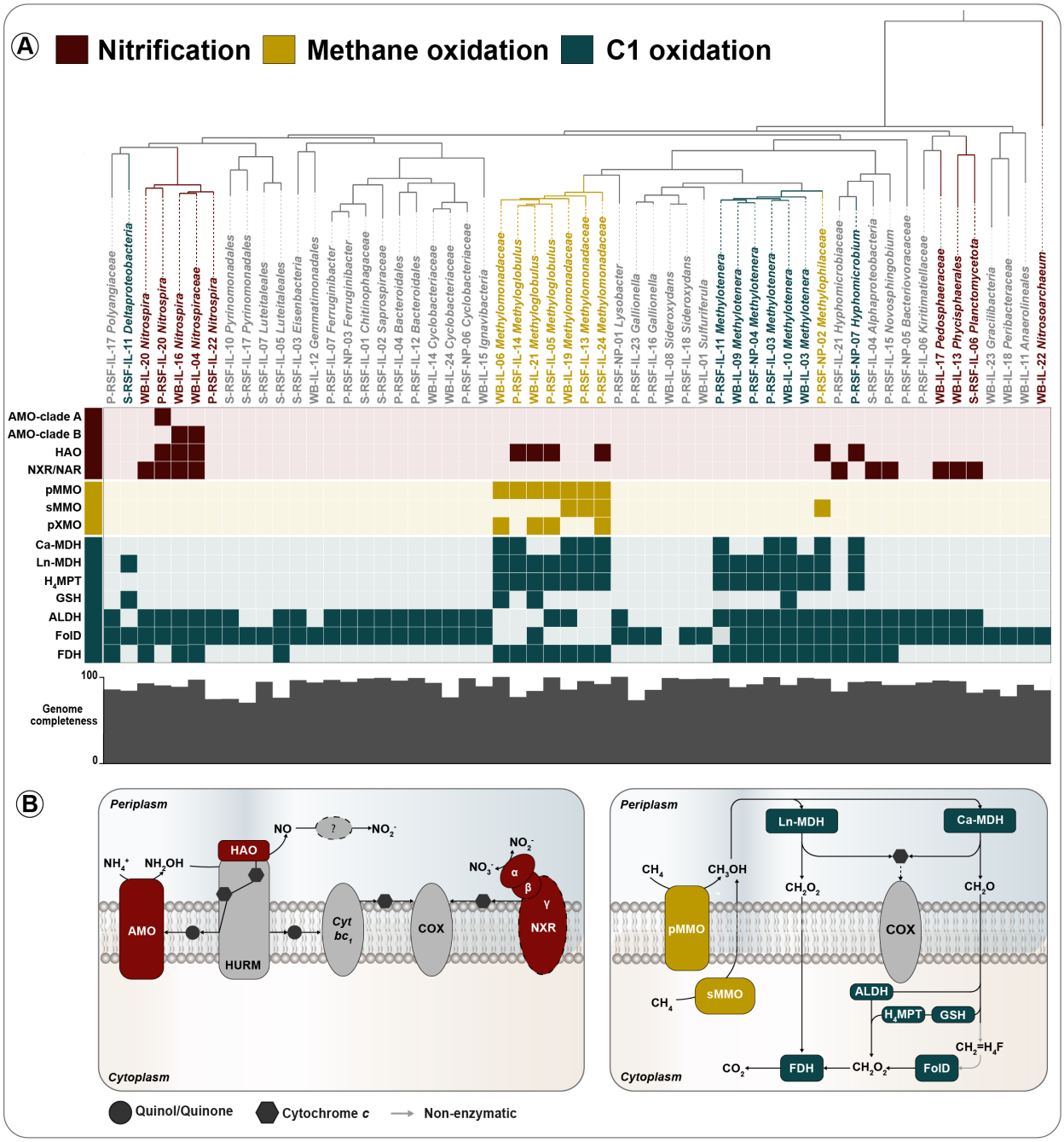


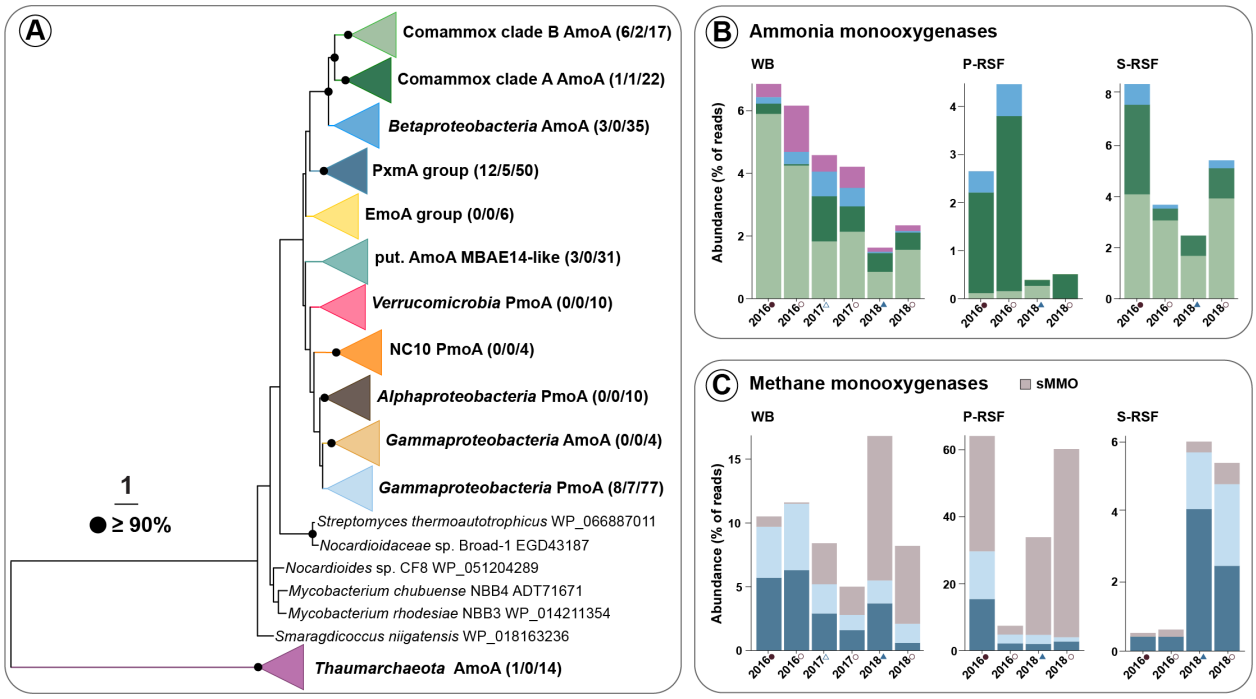
Figure 3. Metabolic potential of the 56 MAGs recovered from DWTP Breehei **A.** Phylogenetic tree based on the concatenated alignment of 49 ribosomal proteins (Table S3) using the anvi′o phylogenomics workflow. Presence or absence of genes for ammonia, nitrite, methane, and C1 utilization are indicated by filled or shaded colored boxes, respectively. Grey bars represent estimated genome completeness. The figure was generated using the anvi′o interactive interface. **B.** Schematic

385 pathway models for complete nitrification, and methane and C1 oxidation, based on enzyme complexes
386 identified in the metagenome. ALDH, aldehyde dehydrogenase; AMO, ammonia monooxygenase;
387 COX, cytochrome *c* oxidase; FDH, formate dehydrogenase; FolD, methylenetetrahydrofolate
388 dehydrogenase/cyclohydrolase; GSH, glutathione-linked formaldehyde oxidation; HAO,
389 hydroxylamine dehydrogenase; H4MPT, tetrahydromethanopterin; HURM, hydroxylamine-
390 ubiquinone redox module; pMMO, particulate methane monooxygenase; sMMO, soluble methane
391 monooxygenase; MDH, lanthanide and calcium-dependent methanol dehydrogenases; NAR, nitrate
392 reductase; NXR, nitrite oxidoreductase.

393 *Ammonia and nitrite oxidation*

394 16S rRNA gene sequence analysis revealed the presence of nitrifying bacteria affiliated with
395 the families *Nitrospiraceae* and *Nitrosomonadaceae* in the DWTP Breehei. *Nitrospiraceae*
396 dominated the nitrifying microbial community in all samples, but their abundance patterns
397 differed along the different sampling locations, with the lowest abundance in P-RSF. In
398 contrast, canonical ammonia-oxidizing bacteria (AOB) affiliated with the *Nitrosomonaceae*
399 showed very low abundance in all samples (Figure S3). Metagenomic consensus binning
400 allowed the recovery of five *Nitrospira* MAGs from WB and P-RSF and one MAG of an
401 ammonia-oxidizing archaeum (AOA) affiliated with the genus *Nitrosoarchaeum*. Despite the
402 detection of *Nitrosomonadaceae* in WB and S-RSF samples in 16S rRNA-based analyses, no
403 high or medium-quality metagenomic bin of this taxonomic group was recovered.
404 All MAGs were screened for key genes of autotrophic nitrification, including the gene sets
405 encoding ammonia monooxygenase (AMO) and nitrite oxidoreductase (NXR). The AMO
406 complex catalyzes the oxidation of ammonia to hydroxylamine (Figure 3B) and belongs to the
407 family of copper-containing membrane monooxygenases (CuMMOs or XMO) (Khadka et al.,
408 2018). Based on the phylogeny of AMO subunit A (AmoA), comammox *Nitrospira* form two
409 monophyletic clades referred to as clade A and B (Daims et al., 2015). AmoA is often used as
410 a functional and phylogenetic marker for ammonia-oxidizing microorganisms (Junier et al.,
411 2008; Pester et al., 2012; Pjevac et al., 2017) and thus AmoA was used to examine the full

412 ammonia oxidation potential in the DWTP Breehei. Phylogenetic analysis placed our
413 metagenome-derived AmoA sequences into five divergent groups (Figure 4A).



414

415 **Figure 4.** Diversity and abundance of ammonia and methane-oxidizing microorganisms in DWTP
416 Breehei. (**A**) Phylogenetic tree of CuMMO subunit A proteins recovered from the DWTP metagenomes
417 by gene-centric and genome-resolved approaches. Numbers in brackets indicate sequences per
418 metagenome, number of recovered MAGs, and references datasets for each clade, respectively. Amo,
419 ammonia monooxygenase; Emo, ethane monooxygenase; Pmo/Pxm, particulate methane
420 monooxygenase. Bootstrap support values ≥90% are indicated by black dots; the scale bar indicates
421 estimated amino acid substitutions**.** Normalized abundances of (**B**) ammonia and (**C**) methane oxidizers
422 are shown as proportion of the recovered ammonia and methane monooxygenases, respectively, to the
423 normalized *rpoB* abundance. The *mmoX* gene was used to calculate the abundance of sMMO-containing
424 methanotrophs; most *mmoX* reads were recruited by MAG P-RSF-NP-02 (70-100%). As most available
425 high-quality betaproteobacterial AOB genomes (n=17, comp. ≥90%, number of contigs <5) contain
426 between 2 and 3 *amoA* copies, and MAG P-RSF-NP-02 contained 4 *mmoX* copies, abundances of both
427 genes were normalized for gene copy numbers. Sample labels correspond to Figure 2.

428 Of the five *Nitrospira* MAGs recovered in this study, one (P-RSF-IL-20) contained *amoA*
429 genes affiliated with clade A, and two (WB-IL-04, WB-IL-16) with clade B (Figure 3A, Table
430 S3). In protein-based phylogenetic analyses, most of the clade A and B comammox AmoA
431 sequences most closely clustered with sequences derived from DWTP metagenomes (Figure

432 S4). Clade A comammox *amoA* genes were detected in all samples from the DWTP Breehei
433 and were the most abundant *amoA* type in P-RSF (0.1-3.6%). In general, ammonia oxidizers
434 were more abundant in 2016 samples than in 2018 (Figure 4B). The clade A comammox
435 *Nitrospira* MAG (P-RSF-IL-20) had the highest coverage in the 2016 P-RSF sample extracted
436 with the Blood & Tissue Kit, while this coverage decreased drastically with another DNA
437 extraction method, indicating a strong extraction bias as discussed above (Table S2).
438 Consequently, the average coverage of this MAG was higher in S-RSF than in P-RSF. In
439 contrast, clade B comammox *amoA* recruited the largest number of reads in WB (0.8-5.9%)
440 and S-RSF samples (1.7-4.1%; Figure 6A), but none of the two clade B-affiliated MAGs was
441 detected in S-RSF, suggesting that not all clade B comammox *Nitrospira* genomes were
442 recovered. The low abundance of clade B comammox *amoA* in P-RSF (Figure 4B) indicates
443 an adaptation of clade B comammox *Nitrospira* to specific niches. In addition to complete
444 nitrifiers, the genus *Nitrospira* also contains canonical nitrite-oxidizing bacteria (NOB) (Daims
445 et al., 2016). Both canonical as well as comammox *Nitrospira* employ the NXR complex for
446 the oxidation of nitrite to nitrate (Figure 3B). Except for one MAG (P-RSF-IL-22), all
447 *Nitrospira* MAGs contained a *nxrAB* gene cluster encoding for the alpha and beta subunit of
448 the NXR complex (Figure 3A, Table S3). MAG P-RSF-IL-22 was of medium quality (74.4%
449 estimated completeness; Table 2) and did not contain any of the genes required for nitrification
450 (Table S3). Consistent with previous studies (Palomo et al., 2019; Poghosyan et al., 2019),
451 phylogenomic analysis using a concatenated alignment of 91 single-copy core genes showed
452 that clade A and B comammox *Nitrospira* formed monophyletic clades within *Nitrospira*
453 lineage II, and all *amoA*-containing MAGs were correctly affiliated with their respective
454 comammox clade (Figure 5A).

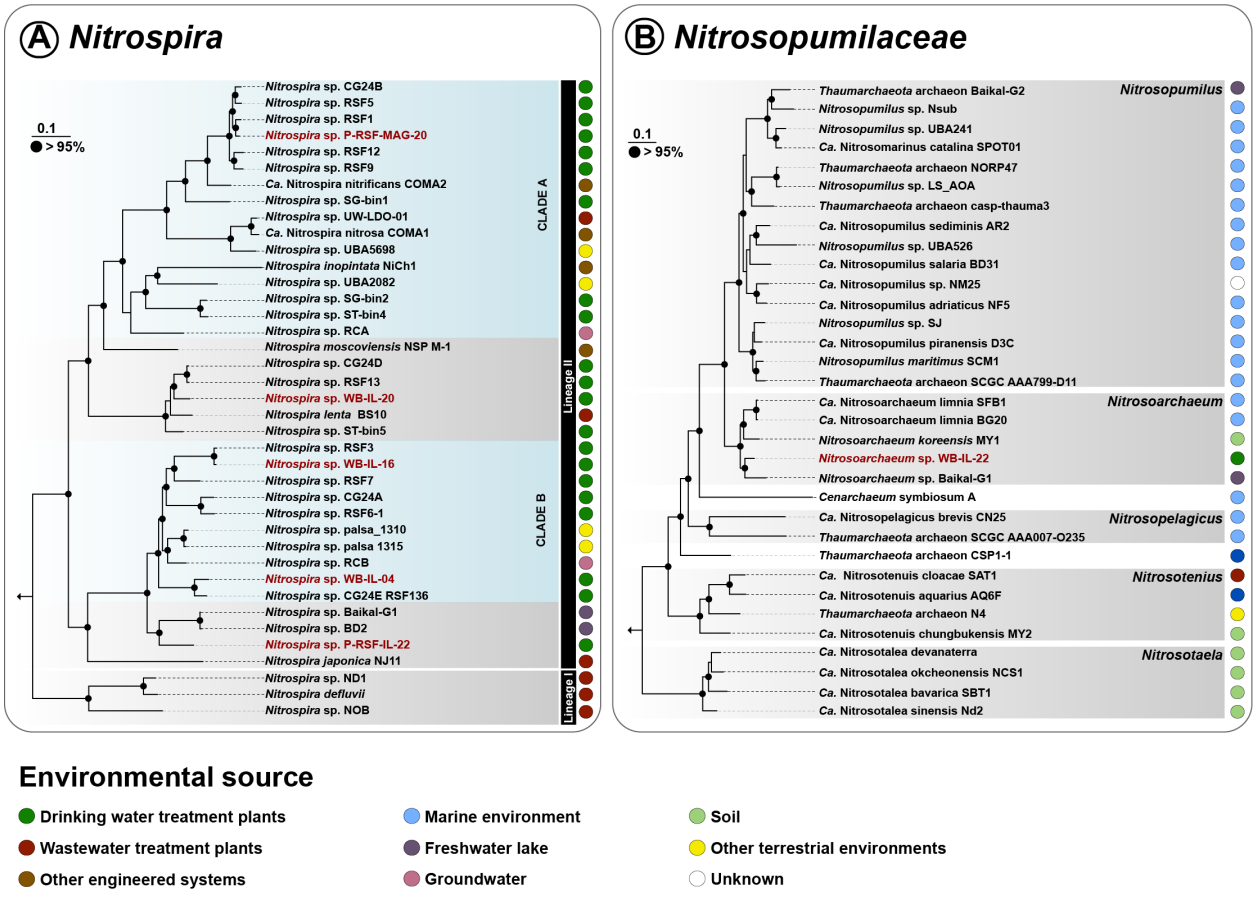
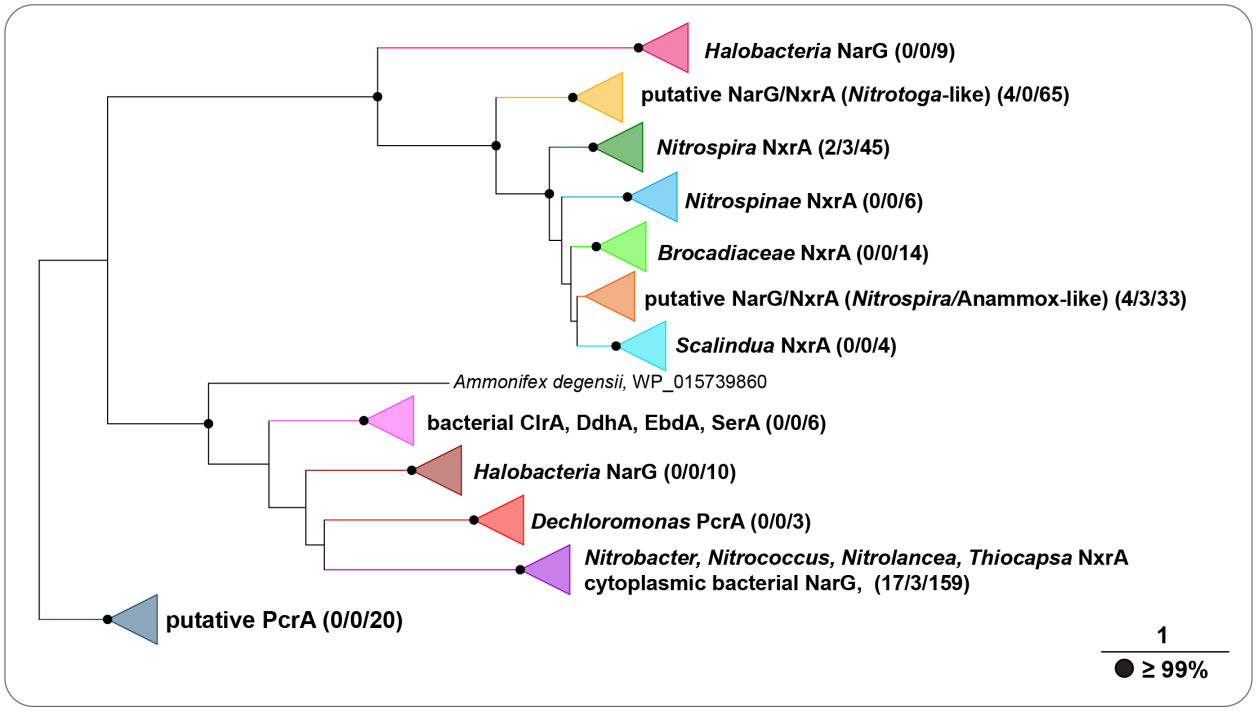


Figure 5. Maximum-likelihood phylogenomic analysis of **(A)** *Nitrospiraceae* based on 91 bacterial single-copy core genes and **(B)** *Nitrosopumilaceae* based on 162 archaeal proteins. DWTP genomes obtained in this study are shown in red. Bootstrap support values ≥95% are indicated by black dots; the scale bars indicate estimated nucleotide **(A)** and amino acid **(B)** substitutions. The positions of the outgroups are indicated by arrows. Environmental origins of the respective genomes are represented by colored circles. Genome names correspond to NCBI RefSeq entries, while in **(B)** classifications at higher taxonomic levels (gray shadings) are according to the GTDB database (Table S4). In **(A)** *Nitrospira* sublineages are delineated by black bars, the comammox clades are highlighted by turquoise shadings.

One MAG (WB-IL-22) affiliated with the genus *Nitrosoarchaeum* (AOA; Figure 5B) was retrieved from the WB metagenome; however, it lacked an *amoA* sequence. Since a contig (<1200 bp) containing a *Nitrosoarchaeum*-like *amoA* was identified in the metagenome, the gene most likely is missing from MAG WB-IL-22 due to the size cutoff used during binning (1500 bp). Consistent with gene-based analyses (0.1-1.7%; Figure 4B) the *Nitrosoarchaeum* MAG was found solely in WB, where it accounted for merely ~0.2% of all the assembled reads

in WB metagenomes (Figure 2, Table S2). The recovered archaeal AmoA sequences had high similarities to *Nitrosarchaeum koreense* and sequences derived from metagenomic analyses of diverse habitats (Figure S4). Betaproteobacterial *amoA* sequences of members of the genera *Nitrosomonas* and *Nitrosospira* were identified in all samples, but were of low abundance. In addition to the characterized ammonia-oxidizing clades, recently the unclassified Gammaproteobacteria "MBAE14" genome was found to contain putative AMO genes (Mori et al., 2019). From the DWTP Breehei metagenomes, three putative AmoA sequences clustered in this novel sequence group despite low similarities to the putative enzyme of MBAE14 (Figure 4A, S4).

The NXR belongs to the type II dimethyl sulfoxide (DMSO) reductase enzyme family (Lücker et al., 2010; Meincke et al., 1992), which also contains respiratory nitrate reductases (NARs) (Simon and Klotz, 2013). Consequently, many HMM profiling approaches fail to differentiate the two homologous groups. Therefore, a phylogenetic tree was constructed to classify the catalytic alpha subunits (NxrA/NarG) identified in the DWTP metagenomes (Figure 6).



486 **Figure 6.** Phylogenetic analysis of NxrA/NarG sequences and related proteins. Numbers in brackets
487 indicate sequences per from DWTP Breehei metagenome, recovered MAGs and reference sequences
488 used for each group. Clr, chlorate reductase; Ddh, dimethylsulfide dehydrogenase; Ebd, ethylbenzene
489 dehydrogenase; Nar, nitrate reductases; Nxr, nitrite oxidoreductase; Pcr, perchlorate reductase; Ser,
490 selenate reductase. Bootstrap support values ≥99% are indicated by black dots; the scale bar indicates
491 estimated amino acid substitutions.

492 Besides the *Nitrospira* genomes, six additional MAGs contained NXR/NAR gene clusters
493 (Figure 3A). Phylogenetic analysis revealed that the sequences derived from
494 alphaproteobacterial MAGs (P-RSF-IL-15, P-RSF-IL-21, S-RSF-IL-04) were closely affiliated
495 with NarG sequences of known nitrate reducers (Figure S5). The sequences from a MAG
496 affiliated with the *Pedosphaeraceae* (WB-IL-17) and two *Planctomycetota* MAGs (WB-IL-
497 13, S-RSF-IL-06) shared low similarities with known NXRs, but clustered with *Nitrospira,*
498 *Nitrospina* and anammox bacteria (Figure S5). While the catalytic NxrA subunit of WB-IL-13
499 contains one Fe-S domain typical for DMSO reductase family molybdoproteins (Lücker et al.,
500 2010), this binding motif for Fe-S clusters is absent from the WB-IL-17 and S-RSF-IL-04
501 NXR-like proteins. Similar to other periplasmic NXRs (Lücker et al., 2013; Lücker et al., 2010;
502 Spieck et al., 1998), these putative NXRs contain N-terminal twin-arginine motifs for protein
503 translocation into the periplasmic space. Putative genes encoding the NXR gamma subunit
504 (*nxrC*) were found to be co-localized with *nxrAB* in the WB-IL-13 and WB-IL-17 MAGs.
505 Similar to *Nitrospina gracilis* (Lücker et al., 2013) and "*Ca.* Nitrotoga fabula" (Kitzinger et
506 al., 2018), these putative NxrCs contain N-terminal signal peptides necessary for translocation
507 via the Sec pathway. The *Pedosphaeraceae* (WB-IL-17) and one of the *Planctomycetota*
508 MAGs that affiliated with the *Phycisphaerales* (WB-IL-13) were present in WB samples, but
509 of low abundance (Figure 2, Table S1). The second, unclassified *Planctomycetota* MAG (S-
510 RSF-IL-06) was detected only in S-RSF and accounted for 0.5-2% of all assembled reads in
511 the S-RSF metagenomes (Table S1). Additionally, some of the metagenome-derived

512 NXR/NAR sequences were placed into the clade containing the novel NXR type of "*Ca.*
513 Nitrotoga fabula" (Figure S4B) (Kitzinger et al., 2018).

514 *Methane and one-carbon metabolism*

515 Taxonomic profiling of the extracted 16S rRNA genes (Figure S3) identified
516 *Gammaproteobacteria* as the most dominant taxa in WB (40-55%) and P-RSF (61-67%)
517 samples. On the family level, *Methylomonadaceae* and *Methylophilaceae* formed the most
518 abundant groups in WB and P-RSF (Figure S3B). In total, 14 MAGs belonging to these two
519 phylogenetic groups were recovered, which were mainly present in WB and P-RSF samples
520 (Figure 2).

521 Phylogenomic analysis of the 14 *Methylomonadaceae* and *Methylophilaceae*-affiliated MAGs
522 using a concatenated alignment of 92 single-copy core genes showed that three of the seven
523 *Methylomonadaceae* MAGs were affiliated with the genus *Methyloglobulus* (Figure 7A). Of
524 the remaining MAGs, two (P-RSF-IL-13 and WB-IL-06) were distantly related to
525 *Methylobacter* and one (WB-IL-19) to *Crenothrix*, while P-RSF-IL-24 clustered separately
526 from the known genera within this family (Figure 7A). Six out of seven *Methylophilaceae*
527 MAGs were affiliated with the genus *Methylotenera,* whereas P-RSF-NP-02 formed a separate
528 branch within the *Methylophilaceae* family (Figure 7B).

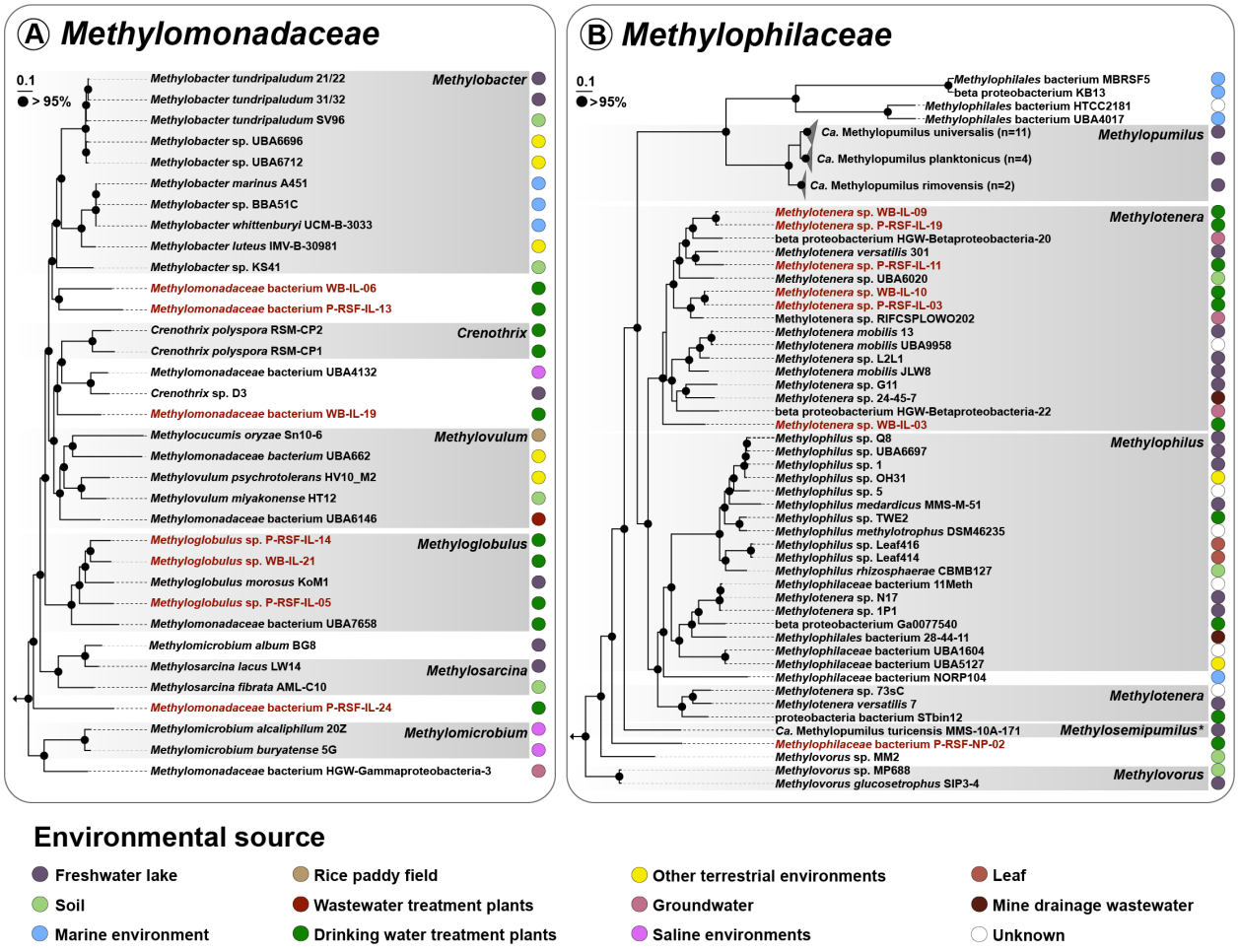


Figure 7. Maximum-likelihood phylogenomic analysis of **(A)** *Methylomonadaceae* and **(B)** *Methylophilaceae* based on 92 bacterial single-copy core genes. MAGs from this study are shown in red. Bootstrap support values ≥95% indicated by black dots; the scale bars indicate estimated nucleotide substitutions. The positions of the outgroups are indicated by arrows. Environmental origins of the respective genomes are represented by colored circles. Genome names correspond to NCBI RefSeq entries, while classifications at higher taxonomic levels are based on GTDB classification (Table S4).

In P-RSF and WB metagenomes, the relative coverage of the *Methylophilaceae* MAGs was much higher than for the *Methylomonadaceae* MAGs (Figure 2, Table S1). These results are consistent with 16S rRNA gene profiling, where *Methylophilaceae* constituted the most abundant family in WB (6-27%) and P-RSF (28-47%) samples and disappeared in S-RSF samples (Figure S3). Particularly, the MAG P-RSF-NP-02 had extremely high coverage in WB (1.6-179×) and P-RSF (27-744×) metagenomes (Table S1). Based on 16S rRNA gene similarity, "*Ca.* Methylosemipumilus turicensis" MMS-10A-171 (Salcher et al., 2019) is the

543 closest described relative of P-RSF-NP-02 (95.7% sequence identity). This value is below the
544 similarity cutoff for species delineation (Stackebrandt and Ebers, 2006), which, in combination
545 with the distinct clustering in the phylogenomic analysis (Figure 7B), indicates that P-RSF-
546 NP-02 probably belongs to a novel genus within the family *Methylophilaceae*. The estimated
547 genomic average nucleotide identity value (ANI; 77.26%) also indicated a novel species
548 distinct from "*Ca.* M. turicensis". The Breehei DWTP metagenomes and the recovered MAGs
549 furthermore were screened for genes encoding the soluble (sMMO) or particulate (pMMO)
550 methane monooxygenases (Table S3). Notably, a complete operon encoding sMMO
551 *(mmoX1X2YBZDC)* was identified in the highly-covered *Methylophilaceae* MAG P-RSF-NP-
552 02 (Figure 3, Table S3). This was surprising as no previously described member of the family
553 *Methylophilaceae* contains any methane monooxygenase (Salcher et al., 2019). P-RSF-NP-02
554 also harbors two additional copies of the *mmoX* gene encoding the alpha subunit of the sMMO
555 (Table S3). The two orphan *mmoX* copies share high sequence identities on amino acid level
556 with the operonal *mmoX1* (99.8-100%), while *mmoX2* has ~97% identity to *mmoX1* and the
557 two orphan *mmoX*. A BLAST search against the NCBI RefSeq database identified that the
558 MmoX copies of P-RSF-NP-02 share 83-84% similarity with the *Methylomicrobium*
559 *buryatense* proteins (Kaluzhnaya et al., 2001). The abundance of methanotrophs in the total
560 community was estimated based on *pmoA, pxmA* and *mmoX* gene coverages (Figure 4C). The
561 sMMO-containing methanotrophs were present at high abundances in the 2018 WB (6-11%)
562 and P-RSF (2.5-56%) samples (Figure 4C), where MAG P-RSF-NP-02 also dominated the
563 total microbial communities (Figure 2). Other samples were dominated by pMMO containing
564 bacteria (Figure 4C.) The gene cluster encoding the structural pMMO subunits *(pmoCAB)* was
565 identified in all *Methylomonadaceae* genomes. Furthermore, five MAGs also harbor the highly
566 divergent *pxmABC* gene cluster encoding for pXMO (Figure 3) (Tavormina et al., 2011).
567 Phylogenetic analysis of the pMMO alpha subunits (PmoA/PxmA) revealed that all sequences

568 recovered in this study, including the metagenome-derived ones, were affiliated with
569 gammaproteobacterial methanotrophs (Figure S4). Besides pMMO, the three
570 *Methylomonadaceae* MAGs (WB-IL-19, P-RSF-IL-13, P-RSF-IL-24) also possess sMMO
571 (Figure 3, Table S3).
572 In both methanotrophs and non-methanotrophic methylotrophs, methanol oxidation to
573 formaldehyde is catalyzed by pyrroloquinoline quinone-dependent (PQQ) methanol
574 dehydrogenases (MDH) (Keltjens et al., 2014). In this study, the rare earth element-dependent
575 MDH (Ln-MDH) was detected in all *Methylomonadaceae and Methylophilaceae*, in one
576 *Deltaproteobacteria* (S-RSF-IL-11), and one *Hyphomicrobium* (P-RSF-NP-07) MAGs. In
577 contrast, the calcium-dependent MDH (Ca-MDH) was found only in some MAGs (Figure 3A,
578 Table S2). Thus, all analyzed methanotrophs and methylotrophs can oxidize methanol either
579 to formaldehyde by Ca-MDH, or directly to formate by Ln-MDH (Figure 3B) (Pol et al., 2014).
580 Four formaldehyde-oxidizing pathways were identified, including oxidation by single
581 aldehyde dehydrogenases, as well as via tetrahydromethanopterin (H_4MPT), tetrahydrofolate
582 (H_4F), or glutathione (GSH)-linked pathways (Figure 3). All the MAGs contained formate
583 dehydrogenases (FDHs), necessary for the oxidation of formate to carbon dioxide.

584 **Discussion**

585 *DWTP microbiome*

586 The DWTP Breehei microbiome was followed over a time period of three years and
587 characterized using genome-resolved and gene-centric metagenomic approaches. The
588 microbial community structure was studied for the filter material of two sequential rapid sand
589 filters and the biofilm formed on the walls of the primary sand filter (Figure 1). Similar to some
590 Danish DWTPs (Gülay et al., 2016), the location within the DWTP had a strong influence on
591 the microbial community composition. In total, 56 dereplicated near-complete MAGs were

592 recovered, comprising 23, 64, and 14% of the total assembled reads for WB, P-RSF, and S-
593 RSF respectively (Table S1). These MAGs expand our knowledge on the genomic inventory
594 of the main microorganisms involved in contaminant removal from groundwater to produce
595 drinking water. The assembly statistics also indicate that the obtained metagenomic
596 information especially for WB and S-RSF only covers a part of the diversity. In general, the
597 genome-centric approach is a powerful tool to analyze the functional potential of an
598 environmental sample based on recovered MAGs. However, due to low abundance or strain
599 diversity, it often is difficult to obtain good-quality genomes for many microorganisms
600 (Sczyrba et al., 2017). Thus, to examine full ammonia and methane oxidation potential, we
601 also used a gene-centric approach (see below).
602 In groundwater, iron exists in the ferrous [Fe(II)] and ferric [Fe(III)] state (Chapelle, 2001) and
603 its oxidation can cause severe operational problems during drinking water production,
604 including bad taste, discoloration, staining, deposition in distribution systems leading to
605 aftergrowth and incidences of high turbidity (Emerson and De Vet, 2015; Sharma et al., 2005).
606 Water quality monitoring at the DWTP Breehei over 16 years indicates 99% iron removal
607 efficiency (Table 1). Under oxic conditions and circumneutral pH, iron oxidation occurs both
608 chemically and biologically in these systems (Tekerlekopoulou et al., 2006). Biological iron
609 oxidation is mediated by chemolithoautotrophic microorganisms that obtain energy from
610 oxidizing ferrous iron (Emerson and De Vet, 2015). However, the absence of an universal
611 metabolic pathway for iron oxidation makes it challenging to identify iron-oxidizing bacteria
612 (Emerson and De Vet, 2015). Here, we obtained four MAGs classified as *Gallionella* and
613 *Sideroxydans* species (Figure 2), which are generally regarded as iron-oxidizing bacteria
614 (Bruun et al., 2010; de Vet et al., 2011; Druschel et al., 2008). These microorganisms were
615 found in WB and P-RSF (Figure 2), where the iron load was high (Table 1). It thus can be

616 assumed that *Gallionellaceae* members are the main drivers of biological iron oxidation in this
617 DWTP.
618 The presence of reduced sulfur compounds such as H_2S, methanethiol, and dimethyl sulfide in
619 anoxic groundwater adds yet another level of complexity to the system (Emerson and De Vet,
620 2015; Sercu et al., 2005). These compounds serve as electron donors for sulfur-oxidizing
621 microorganisms. For sulfide oxidation, sulfide:quinone oxidoreductases (SQR) as well as
622 flavocytochrome *c* sulfide dehydrogenases (FCC) were detected in many of the recovered
623 MAGs (Table S3). However, as sulfide reacts with cytochromes, hemeproteins, and other iron-
624 containing compounds, these enzymes are used for detoxification by many bacteria (Cherney
625 et al., 2010; Marcia et al., 2009; Shahak and Hauska, 2008; Zhang et al., 2013). The MAGs
626 WB-IL-01 and P-RSF-IL-21, classified as *Sulfuriferulla* and *Hyphomicrobium*, respectively,
627 contained the heterodisulfide reductase (HdrCBAHypHdrCB) and sulfur oxidation
628 (SoxXYZAB) enzyme complexes, as well as sulfite oxidase (SoeABC; Table S3). The Sox-
629 Hdr-Soe system is a novel pathway found in some chemolithoautotrophic sulfur-oxidizing
630 bacteria and is involved in volatile organosulfur and inorganic sulfur compound degradation
631 (Koch and Dahl, 2018; Watanabe et al., 2019), and also earlier studies provided evidence that
632 the Hdr complex is involved in sulfur oxidation (Boughanemi et al., 2016; Quatrini et al.,
633 2009). Furthermore, recently it has also been shown in *Hyphomicrobium* that the Hdr complex
634 is essential for oxidation of thiosulfate to sulfate (Koch and Dahl, 2018). The Sox-Hdr-Soe
635 system is found in many *Sulfuriferulla* genomes and SoeABC are suggested to be crucial in
636 sulfur oxidizers that lack SoxCD (Watanabe et al., 2019).
637 Some microbial groups, including *Chloroflexota, Acidobacteriota, Bacteroidota,* and
638 *Myxococcota,* which are originating from groundwater habitats (Griebler and Lueders, 2009;
639 Wegner et al., 2019) are also frequently identified in DWTPs (Albers et al., 2015; Gülay et al.,
640 2016; Palomo et al., 2016). However, we have limited knowledge about their potential role in

641 drinking water treatment. Four MAGs affiliated with *Acidobacteriota* were dominating the
642 microbial community in S-RSF (Figure 2, Figure S3). *Acidobacteriota* are often found in soil
643 under substrate-limited conditions (Hartmann et al., 2015; Jones et al., 2009; Navarrete et al.,
644 2015; Ward et al., 2009). As discussed by Palomo and colleagues (Palomo et al., 2016),
645 *Acidobacteriota* may be involved in carbon cycling in DWTPs. Members of the candidate
646 phyla radiation (CPR) (Brown et al., 2015; Hug et al., 2016) were found only in WB samples
647 (Figure S3). CPR members constitute a substantial fraction of the bacterial diversity (Castelle
648 et al., 2018) and are widely distributed in many environments, including groundwater (Brown
649 et al., 2015; Herrmann et al., 2019; Wegner et al., 2019; Yan et al., 2020) and DWTPs
650 (Bautista-de los Santos et al., 2016; Bruno et al., 2017). One MAG affiliated with "*Ca.*
651 Gracilibacteria" (WB-IL_23) was highly abundant in the WB2018 sample (1-4% of assembled
652 reads) and might have the ability to adapt and survive under low nutrient conditions (Wegner
653 et al., 2019). In groundwater, *Acidobacteriota* and CPR bacteria often co-occur with
654 autotrophic microorganisms (Herrmann et al., 2019). Similar to interactions of nitrifying and
655 heterotrophic bacteria in activated sludge (Okabe et al., 2005), the organic compounds
656 produced by autotrophic bacteria may serve as substrates for *Acidobacteriota* and "*Ca.*
657 Gracilibacteria" in DWTPs.

658 *Ammonia and nitrite metabolism*

659 The removal of ammonium and nitrite is a vital step in drinking water treatment. Although the
660 drinking water produced in DWTP Breehei is of high quality and free of any nitrogen
661 compounds, intermediate nitrite accumulation is observed in the effluent of the P-RSF, which
662 can be caused by incomplete nitrification (de Vet et al., 2012; Wagner et al., 2016). Thus,
663 profound insights into the microbial key players and their metabolic capabilities and limitations
664 are crucial for optimizing and stabilizing N removal in these systems. The genomic potential
665 for nitrification was mainly identified in MAGs affiliated with the genus *Nitrospira,* which

666 however did not display as high abundances as observed in other DWTP systems (Albers et
667 al., 2015; Gülay et al., 2016; Palomo et al., 2016). Comammox *Nitrospira* were the most
668 abundant nitrifying guild in all DWTP Breehei samples (Figure 4B), corroborating previous
669 findings of *Nitrospira* dominating the nitrifying microbial community in DWTPs (Albers et
670 al., 2015; Gülay et al., 2016). Until recently, the dominance of this group was puzzling since
671 *Nitrospira* were always regarded as strict nitrite-oxidizing bacteria. By now, this imbalance in
672 abundance of *Nitrospira* and canonical AOB can be explained by the presence of complete
673 nitrifying *Nitrospira* in many sand filtration systems (Palomo et al., 2018; Pinto et al., 2015;
674 Wang et al., 2017). Notably, complete nitrifiers are the most abundant nitrifying group in
675 several Danish RSFs (Fowler et al., 2018) and were identified as key drivers of ammonia and
676 nitrite oxidation in these systems (Gülay et al., 2019). Consistent with these previous results,
677 comammox *Nitrospira* also apparently outcompeted canonical AOB and AOA in DWTP
678 Breehei. Comammox clade A was found to dominate in P-RSF samples and clade B in WB
679 and S-RSF (Figure 4B). Thus, sampling location substantially influenced the abundance of the
680 two comammox clades, indicating potential niche partitioning between them. The habitat
681 preferences by different nitrifiers might be explained by their ammonia oxidation kinetics and
682 substrate affinities (Kits et al., 2017; Martens-Habbena et al., 2009; Prosser and Nicol, 2012).
683 Based on the kinetic theory of optimal pathway length (Costa et al., 2006) and physiological
684 analyses of *N. inopinata* (Kits et al., 2017), comammox *Nitrospira* are K-strategists that have
685 a competitive advantage in environments with very low ammonium fluxes. However, since no
686 comammox clade B enrichment culture is currently available, we can only speculate about the
687 niche defining metabolic capabilities of this group. A recent study reported that nitrification
688 activity in forest and paddy soils when subjected to ammonium limitation is associated with
689 clade B rather than clade A comammox *Nitrospira* (Wang et al., 2019). Comammox clade B
690 appeared to dominate also in forest soil under increasing nitrogen load and decreasing pH (Shi

691 et al., 2018). Under acidic conditions, ammonia (NH_3) is increasingly protonated to ammonium
692 (NH_4^+), resulting in extremely low concentrations of bioavailable ammonia, the substrate of
693 AMO. Correspondingly, also in Danish RSFs with influent ammonium concentrations ranging
694 from 0.01-0.53 mg-N/L, clade B constituted up to 75% of the total comammox *Nitrospira*
695 population (Fowler et al., 2018). The higher abundance of clade B comammox *Nitrospira* in
696 S-RSF compared to P-RSF observed here also suggests their adaptation to ammonium-depleted
697 environments. These results indicate that clade B comammox *Nitrospira* may exhibit an even
698 lower half-saturation constant (K_s) and higher substrate affinities than clade A species and
699 thrive at extremely low ammonium concentrations. However, it will be required to obtain clade
700 B comammox *Nitrospira* in culture to ascertain the physiology of these enigmatic bacteria.
701 In addition to comammox *Nitrospira*, also two canonical *Nitrospira* MAGs were retrieved from
702 the Breehei DWTP metagenomes, indicating that these nitrite oxidizers can interact with
703 canonical ammonia oxidizers as well as with complete nitrifiers. Moreover, we identified
704 MAGs affiliated with the *Verrucomicrobiota* and *Planctomycetota* phyla (Figure 3) that
705 contained NXR-like sequences with similarity to the enzymes of known NOBs. In phylogenetic
706 analyses of NxrA, they clustered with *Nitrospira*, *Nitrospina,* and anammox bacteria (Figure
707 S4B). The apparent periplasmic orientation of the NxrA and the lack of transmembrane helices
708 in the NxrC subunit suggests that the NXR of these putative NOBs may be soluble, as has also
709 been proposed for *Nitrospina gracilis* (Lücker et al., 2013) and "*Ca.* Nitrotoga fabula"
710 (Kitzinger et al., 2018). However, further studies are needed to analyze the potential nitrite-
711 oxidizing capacity of these bacteria. Although in a very low abundances, *Nitrobacter* species
712 are also detected in some RSFs (Tatari et al., 2017), which we however did not observe in
713 DWTP Breehei.

714 *Methane and one-carbon metabolism*

715 In DWTPs, methane stripping via aeration is preferred over bacterial methane oxidation since
716 microbial activity and growth cause accumulation of extracellular polymeric substances that
717 lead to clogging of the biofilter material (Streese and Stegmann, 2003). However, both WB
718 and P-RSF samples showed high methane-oxidizing capacity (Figure S1). Especially the wall
719 biofilm might counteract methane blowout by oxidizing methane before it leaves the filter and
720 thus reduce methane emissions to the atmosphere. In the global carbon cycle, methane-
721 oxidizing microorganisms play a significant role (Cicerone and Oremland, 1988) as they
722 represent the only known biological methane sink (Aronson et al., 2013). These organisms are
723 able to grow with methane as sole carbon and energy source. The first step of methane
724 oxidation, methane activation and conversion to methanol, is catalyzed either by soluble
725 (sMMO) or particulate (pMMO) methane monooxygenases (Tavormina et al., 2011; Trotsenko
726 and Murrell, 2008). Especially the sMMP is not universal to methanotrophs and only certain
727 phylogenetic groups are known to encode this methane monooxygenase type (Verbeke et al.,
728 2019), usually together with pMMO.

729 Several studies have shown that the majority of the methane-oxidizing bacteria colonizing the
730 granular material of RSFs are affiliated with the gammaproteobacterial *Methylococcaceae*
731 family (Albers et al., 2015; Gülay et al., 2016; Palomo et al., 2016). Recently, this family was
732 reclassified and split into the *Methylomonadaceae, Methylococcaceae,* and
733 *Methylothermaceae* (Parks et al., 2017). Especially members of the *Methylomonadaceae* have
734 been found in many natural and engineered systems (Flynn et al., 2016; Hoefman et al., 2014;
735 Kalyuzhnaya et al., 2015; Kits et al., 2013; Ogiso et al., 2012; Oswald et al., 2017; Parks et al.,
736 2017; Svenning et al., 2011). In this study, all recovered *Methylomonadaceae* MAGs contained
737 pMMO, and some of them additionally encoded sMMO (Figure 3, Table S2). In addition,
738 several methanotrophic MAGs also contained the highly divergent pXMO enzyme complex

739 encoded by the *pxmABC* gene cluster (Tavormina et al., 2011). Previous studies have shown
740 that pXMO is involved in methane oxidation under hypoxic denitrifying conditions in
741 *Methylomonadaceae* strains (Kits et al., 2015a; Kits et al., 2015b), but its exact function in the
742 sand filters remains to be determined.
743 Methane oxidation results in the production of various C1 intermediates including methanol
744 and formate, which can be used as substrates by methylotrophic bacteria. In DWTP Breehei,
745 the P-RSF samples were dominated by members of the *Methylophilaceae* family, which was
746 assumed to accommodate only methylotrophic bacteria incapable of growth on methane.
747 Surprisingly, one of the *Methylophilaceae* MAGs (P-RSF-NP-02) harbored a complete gene
748 operon encoding a sMMO (Figure 3, Table S2), indicating a methanotrophic potential. This
749 finding is substantiated by an earlier study, which demonstrated that a member of
750 methylotrophic genus *Methyloceanibacter* can become methanotrophic by acquiring sMMO
751 (Vekeman et al., 2016). The phylogenomic analysis and low ANI value to other members of
752 this family (77.26%) suggest that P-RSF-NP-02 represents a novel species within the
753 *Methylophilaceae* (Figure 7B) and the extremely high coverage of this MAG indicates a major
754 role in methane removal in the DWTP Breehei. The high abundance of this potential new novel
755 methane-oxidizing bacterium in the P-RSF will be facilitated by the high iron content of the
756 influent water, as the sMMO contains a diiron cluster in the active site (Jasniewski and Que,
757 2018; Wallar and Lipscomb, 1996).

758 **Conclusions**

759 - The metagenomic analyses enabled us to identify key microbial populations involved
760 in the removal of ammonium and methane. The location within the DWTP Breehei was
761 the most influential factor shaping the microbial community.

762 • Clade A comammox *Nitrospira* dominated the nitrifying microbial community in P-
763 RSF, while clade B was most abundant in S-RSF where ammonium concentrations are
764 the lowest and, in the biofilm (WB), which is a predicted niche for comammox bacteria.
765 • The methanotrophic community was dominated by sMMO-containing bacteria,
766 particularly by one novel *Methylophilaceae* member, which might be facilitated by a
767 high iron concentration in the groundwater.

768 **Competing interests**

769 The authors declare that they have no competing interests.

770 **Author contributions**

771 LP and HOdC collected and processed samples. LP and HK analyzed and interpreted the data.
772 JF and GC contributed to bioinformatics analyses. TvA and GC performed Illumina and
773 Nanopore sequencing. HOdC, MJ and MAHJvK were involved in project discussion and data
774 interpretation. SL conceived the research project. LP, HK, and SL wrote the manuscript with
775 input from all the authors.

776 **Data availability**

777 The genome sequences of the 56 MAGs recovered in this study and the raw sequencing data
778 have been deposited in GenBank under BioProject number PRJNA622654.

779 **Acknowledgments**

780 We would like to thank Weren de Vet, Geert Gielens and Kay Bouts for providing necessary
781 information about DWTP Breehei performance and sampling assistance, and Linnea Kop for
782 fruitful discussions. Financial support was provided by the Netherlands Organization for

783 Scientific Research (NWO Talent Programme grants VI.Veni.192.086, 016.Veni.192.062 and
784 016.Vidi.189.050, and Gravitation grants 024.002.001 [NESSC] and 024.002.002 [SIAM]) and
785 the European Research Council (ERC Advanced Grants 339880 [Ecomom] and 669371
786 [VOLCANO]).

787 **References**

788 Albers, C.N., Ellegaard-Jensen, L., Harder, C.B., Rosendahl, S., Knudsen, B.E., Ekelund, F.
789 and Aamand, J. 2015. Groundwater Chemistry Determines the Prokaryotic Community
790 Structure of Waterworks Sand Filters. Environ Sci Technol 49(2), 839-846.

791 Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L.,
792 Loman, N.J., Andersson, A.F. and Quince, C. 2014. Binning metagenomic contigs by
793 coverage and composition. Nat Methods 11(11), 1144-1146.

794 Armenteros, J.J.A., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S.,
795 von Heijne, G. and Nielsen, H. 2019. SignalP 5.0 improves signal peptide predictions
796 using deep neural networks. Nat Biotechnol 37(4), 420-423.

797 Aronson, E.L., Allison, S.D. and Helliker, B.R. 2013. Environmental impacts on the diversity
798 of methane-cycling microbes and their resultant function. Front Microbiol (4:225).

799 Bautista-de los Santos, Q.M., Schroeder, J.L., Sevillano-Rivera, M.C., Sungthong, R., Ijaz,
800 U.Z., Sloan, W.T. and Pinto, A.J. 2016. Emerging investigators series: microbial
801 communities in full-scale drinking water distribution systems - a meta-analysis. Environ
802 Sci-Wat Res Technol 2(4), 631-644.

803 Beech, W.B. and Sunner, J. 2004. Biocorrosion: towards understanding interactions between
804 biofilms and metals. Curr Opin Biotechnol 15(3), 181-186.

805 Boughanemi, S., Lyonnet, J., Infossi, P., Bauzan, M., Kosta, A., Lignon, S., Giudici-Orticoni,
806 M.T. and Guiral, M. 2016. Microbial oxidative sulfur metabolism: biochemical evidence
807 of the membrane-bound heterodisulfide reductase-like complex of the bacterium Aquifex
808 aeolicus. FEMS Microbiol Lett 363(15).

809 Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy,
810 T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N.,
811 Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P.,

812 Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glockner, F.O.,
813 Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S.,
814 Konstantinidis, K.T., Liu, W.T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B.,
815 McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson,
816 G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L.,
817 Banfield, J.F., Hugenholtz, P., Woyke, T. and Genome Stand, C. 2017. Minimum
818 information about a single amplified genome (MISAG) and a metagenome-assembled
819 genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35(8), 725-731.

820 Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J.,
821 Wrighton, K.C., Williams, K.H. and Banfield, J.F. 2015. Unusual biology across a group
822 comprising more than 15% of domain Bacteria. Nature 523(7559), 208-211.

823 Bruno, A., Sandionigi, A., Rizzi, E., Bernasconi, M., Vicario, S., Galimberti, A., Cocuzza, C.,
824 Labra, M. and Casiraghi, M. 2017. Exploring the under-investigated "microbial dark
825 matter" of drinking water treatment plants. Sci Rep 7, 44350.

826 Bruun, A.M., Finster, K., Gunnlaugsson, H.P., Nornberg, P. and Friedrich, M.W. 2010. A
827 Comprehensive Investigation on Iron Cycling in a Freshwater Seep Including Microscopy,
828 Cultivation and Molecular Community Analysis. Geomicrobiol J 27(1), 15-34.

829 Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T.,
830 Soll, D. and Podar, M. 2013. UGA is an additional glycine codon in uncultured SR1
831 bacteria from the human microbiota. Proc Natl Acad Sci USA 110(14), 5540-5545.

832 Camper, A.K. 2004. Involvement of humic substances in regrowth. Int J Food Microbiol
833 92(3), 355-364.

834 Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H. and Banfield, J.F.
835 2018. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and
836 DPANN radiations. Nat Rev Microbiol 16(10), 629-645.

837 Chapelle, F.H. (2001) Ground-water Microbiology and Geochemistry, John Wiley & Sons,
838 New York.

839 Cherney, M.M., Zhang, Y.F., Solomonson, M., Weiner, J.H. and James, M.N.G. 2010. Crystal
840 Structure of Sulfide:Quinone Oxidoreductase from Acidithiobacillus ferrooxidans: Insights
841 into Sulfidotrophic Respiration and Detoxification. J Mol Biol 398(2), 292-305.

842 Cicerone, R.J. and Oremland, R.S. 1988. Biogeochemical aspects of atmospheric methane.
843 Global Biogeochem Cycles 2(4), 299-327.

844 Costa, E., Pérez, J. and Kreft, J.U. 2006. Why is metabolic labour divided in nitrification?
845 Trends Microbiol. 14(5), 213-219.

846 de Vet, W., Dinkla, I.J.T., Rietveld, L.C. and van Loosdrecht, M.C.M. 2011. Biological iron
847 oxidation by Gallionella spp. in drinking water production under fully aerated conditions.
848 Water Res 45(17), 5389-5398.

849 de Vet, W., van Loosdrecht, M.C.M. and Rietveld, L.C. 2012. Phosphorus limitation in
850 nitrifying groundwater filters. Water Res 46(4), 1061-1069.

851 Delmont, T.O. and Eren, A.M. 2016. Identifying contamination with advanced visualization
852 and analysis practices: metagenomic approaches for eukaryotic genome assemblies. PeerJ
853 4, e1839.

854 Druschel, G.K., Emerson, D., Sutka, R., Suchecki, P. and Luther, G.W. 2008. Low-oxygen
855 and chemical kinetic constraints on the geochemical niche of neutrophilic iron(II) oxidizing
856 microorganisms. Geochim Cosmochim Acta 72(14), 3358-3370.

857 EC 2016 Synthesis Report on the Quality of Drinking Water in the Union examining Member
858 States' reports for the 2011-2013 period, foreseen under Article 13(5) of Directive
859 98/83/EC. 2016, https://ec.europa.eu/environment/water/water-drink/reporting_en.html.

860 Eddy, S.R. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7(e1002195).

861 Emerson, D. and De Vet, W. 2015. The Role of FeOB in Engineered Water Ecosystems: A
862 Review. J Am Water Works Assn 107(1), E47-E57.

863 Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L. and Delmont,
864 T.O. 2015. Anvi'o: an advanced analysis and visualization platformfor 'omics data. PeerJ
865 3, e1319.

866 Fish, J.A., Chai, B.L., Wang, Q., Sun, Y.N., Brown, C.T., Tiedje, J.M. and Cole, J.R. 2013.
867 FunGene: the functional gene pipeline and repository. Front Microbiol 4, 291.

868 Flynn, J.D., Hirayama, H., Sakai, Y., Dunfield, P.F., Klotz, M.G., Knief, C., Op den Camp,
869 H.J.M., Jetten, M.S.M., Khmelenina, V.N., Trotsenko, Y.A., Murrell, J.C., Semrau, J.D.,
870 Svenning, M.M., Stein, L.Y., Kyrpides, N., Shapiro, N., Woyke, T., Bringel, F.,
871 Vuilleumier, S., DiSpirito, A.A. and Kalyuzhnaya, M.G. 2016. Draft Genome Sequences

872 of Gammaproteobacterial Methanotrophs Isolated from Marine Ecosystems. Genome
873 Announc 4(1), 2.

874 Fowler, S.J., Palomo, A., Dechesne, A., Mines, P.D. and Smets, B.F. 2018. Comammox
875 Nitrospira are abundant ammonia oxidizers in diverse groundwater-fed rapid sand filter
876 communities. Environ Microbiol 20(3), 1002-1015.

877 Griebler, C. and Lueders, T. 2009. Microbial biodiversity in groundwater ecosystems. Freshw
878 Biol 54(4), 649-677.

879 Gülay, A., Fowler, J., Tatari, K., Thamdrup, B., Albrechtsen, H.-J., Al-Soud, W.A., Sørensen,
880 S.J. and Smets, B.F. 2019. DNA and RNA-SIP reveal Nitrospira spp. as key drivers of
881 nitrification in groundwater-fed biofilters. mBio, 10: e01870-01819.

882 Gülay, A.M., S., Albrechtsen, H.J., Abu Al-Soud, W., Sorensen, S.J. and Smets, B.F. 2016.
883 Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed
884 rapid gravity filters. ISME J. 10(9), 2209-2222.

885 Hartmann, M., Frey, B., Mayer, J., Mader, P. and Widmer, F. 2015. Distinct soil microbial
886 diversity under long-term organic and conventional farming. ISME J 9(5), 1177-1194.

887 Herrmann, M., Wegner, C.E., Taubert, M., Geesink, P., Lehmann, K., Yan, L.J., Lehmann, R.,
888 Totsche, K.U. and Kusel, K. 2019. Predominance of Cand. Patescibacteria in Groundwater
889 Is Caused by Their Preferential Mobilization From Soils and Flourishing Under
890 Oligotrophic Conditions. Front Microbiol 10, 1407.

891 Hoefman, S., Heylen, K. and De Vos, P. 2014. Methylomonas lenta sp nov., a methanotroph
892 isolated from manure and a denitrification tank. Int J Syst Evol Microbiol 64, 1210-1217.

893 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J.,
894 Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman,
895 D.A., Finstad, K.M., Amundson, R., Thomas, B.C. and Banfield, J.F. 2016. A new view
896 of the tree of life. Nat Microbiol 1: 16048(5).

897 Hyatt, D. 2010. Prodigal: prokaryotic gene recognition and translation initiation site
898 identification. BMC Bioinf 11(1), 119.

899 Jasniewski, A.J. and Que, L. 2018. Dioxygen Activation by Nonheme Diiron Enzymes:
900 Diverse Dioxygen Adducts, High-Valent Intermediates, and Related Model Complexes.
901 Chem Rev 118(5), 2554-2592.

902 Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R. and Fierer, N. 2009. A
903 comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone
904 library analyses. ISME J 3(4), 442-453.

905 Kalyuzhnaya, M.G., Lamb, A.E., McTaggart, T.L., Oshkin, I.Y., Shapiro, N., Woyke, T. and
906 Chistoserdova, L. 2015. Draft Genome Sequences of Gammaproteobacterial
907 Methanotrophs Isolated from Lake Washington Sediment. Genome Announc 3(2: e00103-
908 15).

909 Khadka, R., Clothier, L., Wang, L., Lim, C.K., Klotz, M.G. and Dunfield, P.F. 2018.
910 Evolutionary History of Copper Membrane Monooxygenases. Front Microbiol 9, 2493.

911 Kim, D., Song, L., Breitwieser, F.P. and Salzberg, S.L. 2016. Centrifuge: rapid and sensitive
912 classification of metagenomic sequences. Genome Res 26(12), 1721-1729.

913 Kits, K.D., Campbell, D.J., Rosana, A.R. and Stein, L.Y. 2015a. Diverse electron sources
914 support denitrification under hypoxia in the obligate methanotroph Methylomicrobium
915 album strain BG8. Front Microbiol 6, 1072.

916 Kits, K.D., Kalyuzhnaya, M.G., Klotz, M.G., Jetten, M.S., Op den Camp, H.J., Vuilleumier,
917 S., Bringel, F., Dispirito, A.A., Murrell, J.C., Bruce, D., Cheng, J.F., Copeland, A.,
918 Goodwin, L., Hauser, L., Lajus, A., Land, M.L., Lapidus, A., Lucas, S., Medigue, C.,
919 Pitluck, S., Woyke, T., Zeytun, A. and Stein, L.Y. 2013. Genome Sequence of the Obligate
920 Gammaproteobacterial Methanotroph Methylomicrobium album Strain BG8. Genome
921 Announc 1(2), e0017013.

922 Kits, K.D., Klotz, M.G. and Stein, L.Y. 2015b. Methane oxidation coupled to nitrate reduction
923 under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp nov type
924 strain FJG1. Environmental Microbiology 17(9), 3219-3232.

925 Kits, K.D., Sedlacek, C.J., Lebedeva, E.V., Han, P., Bulaev, A., Pjevac, P., Daebeler, A.,
926 Romano, S., Albertsen, M., Stein, L.Y., Daims, H. and Wagner, M. 2017. Kinetic analysis
927 of a complete nitrifier reveals an oligotrophic lifestyle. Nature 549(7671), 269-272.

928 Kitzinger, K., Koch, H., Lucker, S., Sedlacek, C.J., Herbold, C., Schwarz, J., Daebeler, A.,
929 Mueller, A.J., Lukumbuzya, M., Romano, S., Leisch, N., Karst, S.M., Kirkegaard, R.,
930 Albertsen, M., Nielsen, P.H., Wagner, M. and Daims, H. 2018. Characterization of the
931 First "Candidatus Nitrotoga" Isolate Reveals Metabolic Versatility and Separate Evolution
932 of Widespread Nitrite-Oxidizing Bacteria. mBio 9(4:e01186-18).

933 Koch, T. and Dahl, C. 2018. A novel bacterial sulfur oxidation pathway provides a new link
934 between the cycles of organic and inorganic sulfur compounds. ISME J 12(10), 2479-2491.

935 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. 2017.
936 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
937 separation. Genome Res 27(5), 722-736.

938 Koster, J. and Rahmann, S. 2012. Snakemake-a scalable bioinformatics workflow engine.
939 Bioinformatics 28(19), 2520-2522.

940 Kowalchuk, G.A., de Bruijn, F.J., Head, I.M., Akkermans, A.D. and van Elsas, J.D. 2004
941 Molecular microbial ecology manual (MMEM), 2nd ed., vol. 1.,
942 Kluwer Academic Publishing, London, United Kingdom.

943 Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. 2001. Predicting
944 transmembrane protein topology with a hidden Markov model: Application to complete
945 genomes. J Mol Biol 305(3), 567-580.

946 Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nat
947 Methods 9(4), 357-359.

948 Li, D.H., Liu, C.M., Luo, R.B., Sadakane, K. and Lam, T.W. 2015. MEGAHIT: an ultra-fast
949 single-node solution for large and complex metagenomics assembly via succinct de Bruijn
950 graph. Bioinformatics 31(10), 1674-1676.

951 Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18),
952 3094-3100.

953 Li, H. and Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler
954 transform. Bioinformatics 26(5), 589-595.

955 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
956 Durbin, R. and Genome Project Data, P. 2009. The Sequence Alignment/Map format and
957 SAMtools. Bioinformatics 25(16), 2078-2079.

958 Li, H.S. and Carlson, K.H. 2014. Distribution and Origin of Groundwater Methane in the
959 Wattenberg Oil and Gas Field of Northern Colorado. Environ Sci Technol 48(3), 1484-
960 1491.

961 Li, Q., Yu, S.L., Li, L., Liu, G.C., Gu, Z.Y., Liu, M.M., Liu, Z.Y., Ye, Y.B., Xia, Q. and Ren,
962 L.M. 2017. Microbial Communities Shaped by Treatment Processes in a Drinking Water

963 Treatment Plant and Their Contribution and Threat to Drinking Water Safety. Front
964 Microbiol 8, 2465.

965 Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai,
966 T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O.,
967 Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff,
968 B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M.,
969 Ludwig, T., Bode, A. and Schleifer, K.H. 2004. ARB: a software environment for
970 sequence data. Nucleic Acids Res 32(4), 1363-1371.

971 Lücker, S., Nowka, B., Rattei, T., Spieck, E. and Daims, H. 2013. The Genome of Nitrospina
972 gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer.
973 Front Microbiol 4, 27.

974 Maksimavičius, E. and Roslev, P. 2020. Methane emission and methanotrophic activity in
975 groundwater fed drinking water treatment plants. Water Supply, ws2020009.

976 Marcia, M., Ermler, U., Peng, G.H. and Michel, H. 2009. The structure of Aquifex aeolicus
977 sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration.
978 Proc. Natl. Acad. Sci. U.S.A. 106(24), 9625-9630.

979 Martens-Habbena, W., Berube, P.M., Urakawa, H., de la Torre, J.R. and Stahl, D.A. 2009.
980 Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria.
981 Nature 461(7266), 976-979.

982 Matsen, F.A., Kodner, R.B. and Armbrust, E.V. 2010. pplacer: linear time maximum-
983 likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree.
984 Bmc Bioinf 11(538).

985 Miller, M.A., Pfeiffer, W. and Schwartz, T. 2010 Creating the CIPRES Science Gateway for
986 inference of large phylogenetic trees. Gateway Computing Environments Workshop
987 (GCE), pp. 1-8.

988 Na, S.I., Kim, Y.O., Yoon, S.H., Ha, S.M., Baek, I. and Chun, J. 2018. UBCG: Up-to-date
989 bacterial core gene set and pipeline for phylogenomic tree reconstruction. J Microbiol
990 56(4), 280-285.

991 Navarrete, A.A., Venturini, A.M., Meyer, K.M., Klein, A.M., Tiedje, J.M., Bohannan, B.J.M.,
992 Nusslein, K., Tsai, S.M. and Rodrigues, J.L.M. 2015. Differential Response of

993 Acidobacteria Subgroups to Forest-to-Pasture Conversion and Their Biogeographic
994 Patterns in the Western Brazilian Amazon. Front Microbiol 6, 1443.

995 Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P.A. 2017. metaSPAdes: a new
996 versatile metagenomic assembler. Genome Res 27(5), 824-834.

997 Ogiso, T., Ueno, C., Dianou, D., Huy, T.V., Katayama, A., Kimura, M. and Asakawa, S. 2012.
998 Methylomonas koyamae sp nov., a type I methane-oxidizing bacterium from floodwater of
999 a rice paddy field. Int J Syst Evol Microbiol 62, 1832-1837.

1000 Oh, S., Hammes, F. and Liu, W.T. 2018. Metagenomic characterization of biofilter microbial
1001 communities in a full-scale drinking water treatment plant. Water Res 128, 278-285.

1002 Okabe, S., Kindaichi, T. and Ito, T. 2005. Fate of 14C-Labeled Microbial Products Derived
1003 from Nitrifying Bacteria in Autotrophic Nitrifying Biofilms. Appl Environ Microbiol
1004 71(7), 3987.

1005 Okoniewska, E., Lach, J., Kacprzak, M. and Neczaj, E. 2007. The removal of manganese,
1006 iron and ammonium nitrogen on impregnated activated carbon. Desalination 206(1-3), 251-
1007 258.

1008 Olm, M.R., Brown, C.T., Brooks, B. and Banfield, J.F. 2017. dRep: a tool for fast and accurate
1009 genomic comparisons that enables improved genome recovery from metagenomes through
1010 de-replication. ISME J 11(12), 2864-2868.

1011 Osborn, S.G., Vengosh, A., Warner, N.R. and Jackson, R.B. 2011. Methane contamination of
1012 drinking water accompanying gas-well drilling and hydraulic fracturing. Proc. Natl. Acad.
1013 Sci. U.S.A. 108(20), 8172-8176.

1014 Oswald, K., Graf, J.S., Littmann, S., Tienken, D., Brand, A., Wehrli, B., Albertsen, M., Daims,
1015 H., Wagner, M., Kuypers, M.M.M., Schubert, C.J. and Milucka, J. 2017. Crenothrix are
1016 major methane consumers in stratified lakes. ISME J. 11(9), 2124-2140.

1017 Palomo, A., Dechesne, A. and Smets, B.F. 2019. Genomic profiling of Nitrospira species
1018 reveals ecological success of comammox Nitrospira. bioRxiv, 612226.

1019 Palomo, A., Jane Fowler, S., Gulay, A., Rasmussen, S., Sicheritz-Ponten, T. and Smets, B.F.
1020 2016. Metagenomic analysis of rapid gravity sand filter microbial communities suggests
1021 novel physiology of Nitrospira spp. ISME J 10(11), 2569-2581.

1022 Palomo, A., Pedersen, A.G., Fowler, S.J., Dechesne, A., Sicheritz-Ponten, T. and Smets, B.F.
1023 2018. Comparative genomics sheds light on niche differentiation and the evolutionary
1024 history of comammox Nitrospira. ISME J. 12(7), 1779-1793.

1025 Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A. and
1026 Hugenholtz, P. 2018. A standardized bacterial taxonomy based on genome phylogeny
1027 substantially revises the tree of life. Nature Biotechnology 36(10), 996-1004.

1028 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W. 2015. CheckM:
1029 assessing the quality of microbial genomes recovered from isolates, single cells, and
1030 metagenomes. Genome research 25(7), 1043-1055.

1031 Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B., Evans, P.N.,
1032 Hugenholtz, P. and Tyson, G.W. 2017. Recovery of nearly 8,000 metagenome-assembled
1033 genomes substantially expands the tree of life. Nat Microbiol 3(2), 253-253.

1034 Pinto, A.J., Marcus, D.N., Ijaz, U.Z., Santos, Q., Dick, G.J. and Raskin, L. 2015. Metagenomic
1035 Evidence for the Presence of Comammox Nitrospira-Like Bacteria in a Drinking Water
1036 System. Msphere 1(1), e00054-00015.

1037 Pinto, A.J., Xi, C.W. and Raskin, L. 2012. Bacterial Community Structure in the Drinking
1038 Water Microbiome Is Governed by Filtration Processes. Environ Sci Technol 46(16), 8851-
1039 8859.

1040 Poghosyan, L., Koch, H., Lavy, A., Frank, J., van Kessel, M.A.H.J., Jetten, M.S.M., Banfield,
1041 J.F. and Lücker, S. 2019. Metagenomic recovery of two distinct comammox Nitrospira
1042 from the terrestrial subsurface. Environ Microbiol 21(10), 3627-3637.

1043 Pol, A., Barends, T.R.M., Dietl, A., Khadem, A.F., Eygensteyn, J., Jetten, M.S.M. and Op den
1044 Camp, H.J.M. 2014. Rare earth metals are essential for methanotrophic life in volcanic
1045 mudpots. Environ Microbiol 16(1), 255-264.

1046 Proctor, C.R. and Hammes, F. 2015. Drinking water microbiology - from measurement to
1047 management. Curr Opin Biotechnol 33, 87-94.

1048 Prosser, J.I. and Nicol, G.W. 2012. Archaeal and bacterial ammonia-oxidisers in soil: the
1049 quest for niche specialisation and differentiation. Trends Microbiol 20(11), 523-531.

1050 Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D.S. and Bonnefoy, V. 2009.
1051 Extending the models for iron and sulfur oxidation in the extreme Acidophile
1052 Acidithiobacillus ferrooxidans. Bmc Genomics 10(394).

1053 Racine, J.S. 2012. RStudio: A Platform-Independent IDE for R and Sweave. J Appl Econ
1054 27(1), 167-172.

1055 Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A.,
1056 Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert,
1057 S.M., Liu, W.T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M.,
1058 Hugenholtz, P. and Woyke, T. 2013. Insights into the phylogeny and coding potential of
1059 microbial dark matter. Nature 499(7459), 431-437.

1060 Rittmann, B.E., Tangy, Y., Meyer, K. and Bellamy, W.D. 2012 Biological processes, chapter
1061 17. In: Water treatment design. Ed Randtke SJ and Horsley

1062 MB, Am Water Works Assoc (AWWA), McGraw Hill.

1063 Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda,
1064 S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T.S.,
1065 Shapiro, N., Blood, P.D., Gurevich, A., Bai, Y., Turaev, D., DeMaere, M.Z., Chikhi, R.,
1066 Nagarajan, N., Quince, C., Meyer, F., Balvočiūtė, M., Hansen, L.H., Sørensen, S.J., Chia,
1067 B.K.H., Denis, B., Froula, J.L., Wang, Z., Egan, R., Don Kang, D., Cook, J.J., Deltel, C.,
1068 Beckstette, M., Lemaitre, C., Peterlongo, P., Rizk, G., Lavenier, D., Wu, Y.-W., Singer,
1069 S.W., Jain, C., Strous, M., Klingenberg, H., Meinicke, P., Barton, M.D., Lingner, T., Lin,
1070 H.-H., Liao, Y.-C., Silva, G.G.Z., Cuevas, D.A., Edwards, R.A., Saha, S., Piro, V.C.,
1071 Renard, B.Y., Pop, M., Klenk, H.-P., Göker, M., Kyrpides, N.C., Woyke, T., Vorholt, J.A.,
1072 Schulze-Lefert, P., Rubin, E.M., Darling, A.E., Rattei, T. and McHardy, A.C. 2017.
1073 Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics
1074 software. Nat Methods 14(11), 1063-1071.

1075 Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14),
1076 2068-2069.

1077 Sercu, B., Nunez, D., Van Langenhove, H., Aroca, G. and Verstraete, W. 2005. Operational
1078 and microbiological aspects of a bioaugmented two-stage biotrickling filter removing
1079 hydrogen sulfide and dimethyl sulfide. Biotechnol Bioeng 90(2), 259-269.

1080 Shahak, Y. and Hauska, G. (2008) Sulfur Metabolism in Phototrophic Organisms. Hell, R.,
1081 Dahl, C., Knaff, D. and Leustek, T. (eds), pp. 319-335, Springer Netherlands, Dordrecht.

1082 Sharma, S.K., Petrusevski, B. and Schippers, J.C. 2005. Biological iron removal from
1083 groundwater: a review. J. Water Supply Res. Technol. AQUA 54(4), 239-247.

1084 Shi, X.Z., Hu, H.W., Wang, J.Q., He, J.Z., Zheng, C.Y., Wan, X.H. and Huang, Z.Q. 2018.
1085 Niche separation of comammox Nitrospira and canonical ammonia oxidizers in an acidic
1086 subtropical forest soil under long-term nitrogen deposition. Soil Biol Biochem 126, 114-
1087 122.

1088 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1089 large phylogenies. Bioinformatics 30(9), 1312-1313.

1090 Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A.,
1091 Midgley, P.M., Bex, V. and Xia, Y. 2013 IPCC, 2013: Climate Change 2013: The Physical
1092 Science Basis. Working Group I Contribution to the IPCC 5th Assessment Report, IPCC,
1093 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535
1094 pp.

1095 Streese, J. and Stegmann, R. 2003. Microbial oxidation of methane from old landfills in
1096 biofilters. Waste Manage 23(7), 573-580.

1097 Svenning, M.M., Hestnes, A.G., Wartiainen, I., Stein, L.Y., Klotz, M.G., Kalyuzhnaya, M.G.,
1098 Spang, A., Bringel, F., Vuilleumier, S., Lajus, A., Medigue, C., Bruce, D.C., Cheng, J.F.,
1099 Goodwin, L., Ivanova, N., Han, J., Han, C.S., Hauser, L., Held, B., Land, M.L., Lapidus,
1100 A., Lucas, S., Nolan, M., Pitluck, S. and Woyke, T. 2011. Genome Sequence of the Arctic
1101 Methanotroph Methylobacter tundripaludum SV96. J Bacteriol 193(22), 6418-6419.

1102 Tatari, K., Musovic, S., Gulay, A., Dechesne, A., Albrechtsen, H.J. and Smets, B.F. 2017.
1103 Density and distribution of nitrifying guilds in rapid sand filters for drinking water
1104 production: Dominance of Nitrospira spp. Water Res 127, 239-248.

1105 Tatusov, R.L., Koonin, E.V. and Lipman, D.J. 1997. A genomic perspective on protein
1106 families. Science 278(5338), 631-637.

1107 Tavormina, P.L., Orphan, V.J., Kalyuzhnaya, M.G., Jetten, M.S.M. and Klotz, M.G. 2011. A
1108 novel family of functional operons encoding methane/ammonia monooxygenase-related
1109 proteins in gammaproteobacterial methanotrophs. Environ Microbiol Rep 3(1), 91-100.

1110 Tekerlekopoulou, A.G., Vasiliadou, I.A. and Vayenas, D.V. 2006. Physico-chemical and
1111 biological iron removal from potable water. Biochem Eng J 31(1), 74-83.

1112 Trifinopoulos, J., Nguyen, L.T., von Haeseler, A. and Minh, B.Q. 2016. W-IQ-TREE: a fast
1113 online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44(W1),
1114 W232-W235.

1115 Trussell, R.R., Hand, D.W., Tchobanoglous, G., Howe, K.J. and Crittenden, J.C. (2012)
1116 Principles of Water Treatment, John Wiley & Sons Inc, Hoboken, New Jersey.

1117 van der Wielen, P.W.J.J., Voost, S. and van der Kooij, D. 2009. Ammonia-Oxidizing Bacteria
1118 and Archaea in Groundwater Treatment and Drinking Water Distribution Systems. Appl
1119 Environ Microbiol 75(14), 4687-4695.

1120 Vaser, R., Sovic, I., Nagarajan, N. and Sikic, M. 2017. Fast and accurate de novo genome
1121 assembly from long uncorrected reads. Genome Res 27(5), 737-746.

1122 Vekeman, B., Kerckhof, F.M., Cremers, G., de Vos, P., Vandamme, P., Boon, N., Op den
1123 Camp, H.J.M. and Heylen, K. 2016. New Methyloceanibacter diversity from North Sea
1124 sediments includes methanotroph containing solely the soluble methane monooxygenase.
1125 Environ Microbiol 18(12), 4523-4536.

1126 Wagner, F.B., Nielsen, P.B., Boe-Hansen, R. and Albrechtsen, H.J. 2016. Copper deficiency
1127 can limit nitrification in biological rapid sand filters for drinking water production. Water
1128 Res 95, 280-288.

1129 Wallar, B.J. and Lipscomb, J.D. 1996. Dioxygen activation by enzymes containing binuclear
1130 non-heme iron clusters. Chem Rev 96(7), 2625-2657.

1131 Wang, Y., Ma, L., Mao, Y., Jiang, X., Xia, Y., Yu, K., Li, B. and Zhang, T. 2017. Comammox
1132 in drinking water systems. Water Res 116, 332-341.

1133 Wang, Z.H., Cao, Y.Q., Zhu-Barker, X., Nicol, G.W., Wright, A.L., Jia, Z.J. and Jiang, X.J.
1134 2019. Comammox Nitrospira clade B contributes to nitrification in soil. Soil Biol Biochem
1135 135, 392-395.

1136 Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G.,
1137 Haft, D.H., Sait, M., Badger, J., Barabote, R.D., Bradley, B., Brettin, T.S., Brinkac, L.M.,
1138 Bruce, D., Creasy, T., Daugherty, S.C., Davidsen, T.M., Deboy, R.T., Detter, J.C., Dodson,
1139 R.J., Durkin, A.S., Ganapathy, A., Gwinn-Giglio, M., Han, C.S., Khouri, H., Kiss, H.,
1140 Kothari, S.P., Madupu, R., Nelson, K.E., Nelson, W.C., Paulsen, I., Penn, K., Ren, Q.H.,
1141 Rosovitz, M.J., Selengut, J.D., Shrivastava, S., Sullivan, S.A., Tapia, R., Thompson, L.S.,
1142 Watkins, K.L., Yang, Q., Yu, C.H., Zafar, N., Zhou, L.W. and Kuske, C.R. 2009. Three
1143 Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These
1144 Microorganisms in Soils. Appl Environ Microbiol 75(7), 2046-2056.

1145 Watanabe, T., Kojima, H., Umezawa, K., Hori, C., Takasuka, T.E., Kato, Y. and Fukui, M.
1146 2019. Genomes of Neutrophilic Sulfur-Oxidizing Chemolithoautotrophs Representing 9
1147 Proteobacterial Species From 8 Genera. Front Microbiol 10(316).

1148 Wegner, C.E., Gaspar, M., Geesink, P., Herrmann, M., Marz, M. and Kusel, K. 2019.
1149 Biogeochemical Regimes in Shallow Aquifers Reflect the Metabolic Coupling of the
1150 Elements Nitrogen, Sulfur, and Carbon. Appl Environ Microbiol 85(9), 0099-2240.

1151 Whitman, W.B., Oren, A., Chuvochina, M., da Costa, M.S., Garrity, G.M., Rainey, F.A.,
1152 Rossello-Mora, R., Schink, B., Sutcliffe, I., Trujillo, M.E. and Ventura, S. 2018. Proposal
1153 of the suffix -ota to denote phyla. Addendum to 'Proposal to include the rank of phylum in
1154 the International Code of Nomenclature of Prokaryotes'. Int. J. Syst. Evol. Microbiol. 68(3),
1155 967-969.

1156 Wickham, H. 2016 ggplot2: Elegant Graphics for Data Analysis, (Springer, Heidelberg),
1157 Springer, Heidelberg.

1158 Wilczak, A., Jacangelo, J.G., Marcinko, J.P., Odell, L.H., Kirmeyer, G.J. and Wolfe, R.L.
1159 1996. Occurrence of nitrification in chloraminated distribution systems. J Am Water
1160 Works Assoc 88(7), 74-85.

1161 Yan, L., Herrmann, M., Kampe, B., Lehmann, R., Totsche, K.U. and Küsel, K. 2020.
1162 Environmental selection shapes the formation of near-surface groundwater microbiomes.
1163 Water Res 170, 115341.

1164 Yang, J.X., Ma, J., Song, D., Zhai, X.D. and Kong, X.J. 2016. Impact of preozonation on the
1165 bioactivity and biodiversity of subsequent biofilters under low temperature conditions-A
1166 pilot study. Front Environ Sci Eng 10(4).

1167 Zhang, Y.F., Yang, Y., Liu, J.S. and Qiu, G.Z. 2013. Isolation and characterization of
1168 Acidithiobacillus ferrooxidans strain QXS-1 capable of unusual ferrous iron and sulfur
1169 utilization. Hydrometallurgy 136, 51-57.

1170 Zhou, J.Z., Bruns, M.A. and Tiedje, J.M. 1996. DNA recovery from soils of diverse
1171 composition. Appl Environ Microbiol 62(2), 316-322.

1172