Metabolic diversity and aero-tolerance in anammox bacteria from geochemically distinct aquifers

Authors: Olivia E. Mosley¹, Emilie Gios¹, Louise Weaver², Murray Close², Chris Daughney³, Rob van der Raaij⁴, Heather Martindale^{4†} and Kim M. Handley^{1*}

¹School of Biological Sciences, The University of Auckland, Auckland, New Zealand

²Institute of Environmental Science and Research, Christchurch, New Zealand

³National Institute of Water and Atmospheric Research, Wellington, New Zealand

⁴GNS Science, Lower Hutt, New Zealand

*Corresponding author: K. M. Handley, kim.handley@auckland.ac.nz

[†]Current address: Ministry for the Environment, Wellington, New Zealand

Abstract

Background: Anaerobic ammonium oxidation (anammox) is important for converting bioavailable nitrogen into dinitrogen gas, particularly in carbon poor environments. Yet, the diversity and prevalence of anammox bacteria in the terrestrial subsurface – a typically oligotrophic environment – is little understood across different geochemical conditions. To determine the distribution and activity of anammox bacteria across a range of aquifer lithologies and physicochemistries, we analysed 16S rRNA genes, metagenomes and metatranscriptomes, and quantified hydrazine synthase genes and transcripts sampled from 59 groundwater wells distributed over 1 240 km².

Results: Data indicate that anammox-associated bacteria (class *Brocadia*) and the anammox process are prevalent in aquifers (identified in aquifers with sandy-gravel, sand-silt and volcanic lithologies). While *Brocadia* diversity decreased with increasing DO, *Brocadia* 16S rRNA genes and hydrazine synthase genes and transcripts (hydrazine synthase, *hzsB*) were detected across a wide range of bulk groundwater dissolved oxygen (DO) concentrations (0 – 10 mg/L). Anammox genes and transcripts (*hzsB*) correlated significantly with those involved in bacterial and archaeal ammonia oxidation (ammonia monooxygenase, *amoA*), which could represent a major source of nitrite for anammox. Differences in anammox community composition were strongly associated with DO and bore depth (and to a lesser extent pH and phosphate), revealing niche differentiation among anammox bacteria in groundwater that was largely driven by water oxygen contents, and not ammonium/nitrite. Eight *Brocadia* genomes (63-95% estimated completeness) reconstructed from a subset of groundwater sites belong to 2 uncharacterized families and 6 novel species (based on average nucleotide identity). Distinct groups of these genomes dominated the anammox-associated community at dysoxic and oxic sites, further reflecting the influence of DO on *Brocadia* composition. Six of the genomes (dominating dysoxic or oxic sites) have genes characteristic of anammox (hydrazine synthase and/or dehydrogenase). These genes, in addition to

aerotolerance genes, belonging to four *Brocadia* genomes, were transcriptionally active, although transcript numbers clearly highest in dyoxic groundwater.

Conclusions: Our findings indicate anammox bacteria contribute to loss of fixed N across diverse anoxic-to-oxic aquifer conditions, and that this is likely supported by nitrite from aerobic ammonia oxidation. Results provide an insight into the distribution and activity of anammox bacteria across distinct aquifer physicochemisties.

Keywords

Aquifer; groundwater; anammox; *Brocadia*; ammonia-oxidizers; aero-tolerance

Introduction

For decades, loss of fixed nitrogen (N) from freshwater environments, including aquifers was exclusively attributed to heterotrophic denitrification [1]. Ammonium (NH_4^+) was deemed inert under anoxic conditions [2] and denitrification was considered to be the dominant anoxic sink for inorganic nitrogen [1]. This view was altered when anaerobic ammonium oxidation (anammox), mediated by autotrophic bacteria from a deep-branching class within the phylum *Planctomycetes*, were discovered [3]. Currently, only 10 anammox species have been enriched in culture [4]; none are axenic. Their physiological characteristics and niches are defined using Monod model parameters [5]. Anammox bacteria oxidize ammonium with nitrite (NO_2^-) to N_2 gas without nitrous oxide emissions [6], and are considered chemoautotrophic [7]. Therefore they tend to outcompete heterotrophic denitrifiers in low carbon systems [8]. However, enrichment experiments have demonstrated that some anammox bacteria have the ability to oxidize small organic acids, such as propionate and acetate, with $\text{NO}_3^-/\text{NO}_2^-$ [9], pointing to greater metabolic versatility than previously assumed for these chemolithoautotrophic organisms.

A growing body of research now shows that anammox bacteria play a significant role in the global aquatic N cycle, based on their discovery in diverse aquatic environments, including oceanic oxygen-minimum zones [10,11], meromictic lakes [12], marine subsurface sediments [13], uncontaminated suboxic groundwater [8] and contaminated groundwaters [14–16]. Of these, aquifers harbour the largest freshwater store on Earth, and can be characterized by long groundwater residence times [17], suitable for slow-growing anammox bacteria (with doubling time of 7–22 days) [18]. Moreover, while groundwater often contains low levels of naturally occurring N species, such as nitrate (<0.25 mg/L) [19] and ammonia (<0.2 mg/L) [20], and frequent inputs of N from agricultural sources [21], some aquifers contain little organic carbon [22], potentially limiting heterotrophic denitrification [1]. Aquifers, as oligotrophic environments [22], may therefore represent ideal ecosystems for chemolithoautotrophic anammox bacteria.

Anammox relies on the presence of both oxidized and reduced inorganic N compounds, and subsequently, is highly active in redox transition zones, where N-species fluctuate between oxidation and reduction reactions [23]. Anammox bacteria are therefore found naturally occurring at the interface of aerobic-anaerobic conditions where NO_2^- containing water and anaerobic water with NH_4^+ mix [24]. N-fluctuation frequently occurs in groundwater due to interactions with surface water, or oxygen penetration from above in unconfined shallow aquifers [25]. Elevated groundwater tables, which promote oxygenation through interactions with the overlying soil and groundwater recharge, stimulate a variety of anammox bacteria[16]. Surface water contaminated with nitrogen-species can be a source of NH_4^+ in aquifers. This ammonium can be partially absorbed by clay minerals [26] or oxidised to NO_2^- or NO_3^- by nitrification [1]. NO_2^- produced by aerobic ammonia-oxidizers, can serve as an electron acceptor in anammox [27]. While anammox bacteria have been observed previously in groundwaters [8], their prevalence and genetic diversity is unclear. Nor is it known how they are impacted by the heterogeneous chemical conditions found in aquifers.

To provide insights into anammox ecology and activity across a diverse range of groundwater ecosystems, we collected groundwater from 59 sites over 4 geographic regions across New Zealand. To the best of our knowledge, this is the largest survey of anammox communities in groundwater to date, encompassing a range of characteristics including variations in nitrogen-species, organic carbon and dissolved oxygen concentrations. By reconstruction and metabolic characterisation of diverse novel *Brocadia* genomes, and by quantifying the distribution, abundance and activity of anammox bacteria across distinct groundwater conditions, we reveal the diverse ecological niches of these bacteria in groundwater. Results provide evidence in support of anammox as an important biological process in aquifers and a major N-sink [16].

Materials and methods

Study sites and sample collection

Groundwater was collected from 59 wells spanning 10 aquifers in Waikato, Wellington and Canterbury regions of New Zealand (Supplementary Table S1), from several aquifer lithologies: alluvial sandy gravel (×71 samples), sand-silt (×1), basalt (×1), shellbed (×1), peat (×1) and ignimbrite (×6). Wells were purged (~3–5 bore volumes), then 3–90 L of groundwater or 0.5–15 L of biomass-enriched groundwater were filtered on-site. Biomass-enriched groundwater was collected from 8 sites in Canterbury (Table S1) directly following standard groundwater collection, and low-frequency sonication (2.43 Kw) for 2 minutes to detach biofilms and aquifer particles [28]. Biomass was captured onto mixed cellulose ester membrane filters (1.2 µm pore size pre-filter over a 0.22 µm filter), using a 142 mm stainless steel filter holder (Merck Millipore Ltd, Cork, Ireland). Both filters were immediately submerged in RNA Later (ThermoFisher Scientific, Waltham, MA, USA), transported on dry ice and stored at -80 °C.

Chemical analysis of water samples

Dissolved oxygen (DO), water temperature, pH, oxidation-reduction potential (ORP) and specific conductance (SPC or conductivity) were measured on-site using a flow-through cell and field probes (YSI EXO sonde 2, YSI PRO + and YSI ProDSS, Yellow Springs, Ohio, USA). Samples were categorized by DO concentration as follows: anoxic (0 mg/L), suboxic (<0.3 mg/L), dysoxic (0.3–3 mg/L) and oxic (>3 mg/L) [29]. Unfiltered groundwater was analysed for P, N, C, S, Mn and alkalinity at Hill Laboratories (Hamilton, New Zealand) (Supplementary Information). To assess denitrification/anammox, excess N₂ gas was quantified using methods by Martindale et al. [30]. Briefly, dissolved Ar, Ne and N₂ were measured by a standard curve via a system comprising two detectors, a pulsed discharge helium ionization detector (Valco Instruments D-4-I-SH14-R), a thermal conductivity detector (Shimadzu TCD-2014) and ultra-high purity helium (He) gas. Excess N₂, attributable to denitrification/anammox, was determined by comparison to inert Ar and Ne gas concentrations.

Nucleic acid extraction and sequencing

Nucleic acid extraction. RNA and DNA were extracted using RNeasy PowerSoil Total RNA and DNeasy PowerSoil Pro kits (Qiagen, Valencia, CA, USA), respectively, with nuclease-free glycogen added to aid RNA precipitation (0.1 µg/µL final concentration, Roche diagnostics, Basel, Switzerland). DNA extractions used 0.14–0.89 g of crushed filter (1–47 extractions per sample). For RNA, 2.12–3.90 g was used per extraction (1 extraction per sample). RNA was DNase treated using the TURBO DNA-free™ Kit (“rigorous” protocol; Invitrogen, Carlsbad, CA, USA). DNA removal was verified via 16S rRNA gene amplification (as below, but over 55 cycles) and gel electrophoresis. Replicate DNA extractions were pooled and concentrated using sodium acetate (0.3M final concentration) and ethanol (2x volume) precipitation with 0.1 µg/µL glycogen (Roche diagnostics) via ethanol precipitation. RNA extractions were concentrated using the Zymo RNA Clean and Concentrator-5 Kit (Zymo research, Irvine, CA, USA).

High molecular weight DNA for whole genome shotgun (WGS) sequencing was verified via 1% agarose gel electrophoresis. Nucleic acids were quantified with a Qubit 3.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA) using dsDNA HS and RNA HS assay kits, and quality checked using a NanoPhotometer (Implen, Munich, Germany). RNA was checked using an Agilent BioAnalyzer with RNA 6000 Nano and Pico chips (Integrated Sciences, NSW, Australia). Samples with a RNA Integrity Number ≥6 or DV200 >30% (percentage of RNA fragments above 200 nucleotides) were used to quantify transcripts and for transcriptomics. Between 24 pg and 1.1 µg of total RNA was converted to cDNA using Superscript III Supermix (Invitrogen, Carlsbad, CA, USA).

Metagenome, metatranscriptome and amplicon sequencing. DNA libraries for 15 samples (gwj01–gwj16) were prepared using the TruSeq Nano DNA Kit (Illumina, San Diego, CA, USA), with a targeted insert size of 550 bp, at the Otago Genomics Facility (University of Otago, NZ), except low-yield sample gwj02, which was prepared with the ThruPLEX DNA-seq Kit (Takara Bio USA, Inc., Mountain View, CA, USA). 2 x 250 bp sequencing was performed

using the Illumina HiSeq 2500 V4 platform. RNA libraries were prepared using the Ovation SoLo RNA-Seq System (NuGEN, Redwood City, CA, USA) using custom probes for rRNA depletion at the Otago Genomics Facility, NZ. Custom rRNA probes were designed by the manufacturer using small and large ribosomal subunit sequences reconstructed from the 16 metagenomes with EMIRGE [31] over 40 iterations with clustering at 97% identity and using the SILVA 132 database [32]. Ribosomal sequences generated were used as target sequences in the design of custom AnyDeplete probes using NuGEN's proprietary algorithm. rRNA and genomic DNA were removed as a part of the library preparation kit, using the custom probes and DNase treatment respectively. Paired-end 2x125 bp reads were generated from RNA libraries using the Illumina HiSeq 2500 V4 platform.

PCR amplification of 16S rRNA genes used modified Earth Microbiome Project primers EMP-16S-515'F and EMP-16S-806'R primers [33,34] with Illumina Nextera adapters, and MyTaq HS Red Mix (Bioline, London, UK). PCR conditions are described in Table S2. Amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). Barcoded libraries, prepared according to Illumina's 16S Metagenomic Sequencing Library Preparation manual, were loaded with 10% PhiX, for 2 x 250 bp sequencing via Illumina MiSeq with V2 chemistry (Auckland Genomics, University of Auckland, NZ).

Amplicon processing

Sequences were quality checked using FastQC v0.11.7 [35], and merged using USEARCH v9.0.2132[36]. Sequences were quality filtered using sickle (minimum Phred score ≥ 30 ; length ≥ 200 bp) with another 10 bp of lower quality sequence removed from each end using USEARCH -fastx_truncate [36]. Sequences were dereplicated and clustered at 97% similarity with chimera removal to generate operational taxonomic units (OTUs) using the UCLUST pipeline [37]. OTUs were classified using USEARCH -sintax with the SILVA SSU Ref NR99 database v132 [32]. Non-prokaryotic and singleton sequences were removed before rarefying to 13,393 using QIIME2 v2018.2 [38].

Quantitative PCR of hydrazine synthase and ammonia monooxygenase genes

Droplet Digital PCR (ddPCR) via the Bio-Rad QX200 platform used 20 µl reactions with 10 µl 2x QX200 ddPCR EvaGreen Supermix (Biorad, Hercules, CA, US), 0.4 µM of forward and reverse primer, 1 µl of DNA or cDNA and nuclease-free water. Primers and PCR conditions for hydrazine synthase (*hzsB*_396F/*hzsB*_742R), archaeal ammonia monooxygenase (Arch-*amoA*/Arch-*amoA*R) and bacterial *amoA* (*amoA*1F/*amoA*-modR) genes are described in Table S2. Negative controls used DPEC-treated water. Positive controls used gBlocks dsDNA fragments of *hzsB* and *amoA* genes (Integrated DNA Technologies, Coralville, Iowa, US). Data was analyzed using the QuantaSoft software package (Bio-Rad). Threshold values for positive droplets were set based on the amplitude of negative and positive droplets using the positive control as reference.

Metagenome assembly and genome binning

Adapters were removed from metagenomic reads using Cutadapt [39]. Reads were trimmed with sickle (Phred score ≥ 30 ; length ≥ 80 bp) and quality checked using FastQC v0.11.7 [35]. All samples were individually assembled using SPAdes v3.11.1 [40] (--meta, -k 43,55,77,99,121). Samples from the same well (groundwater $\pm$ biomass-enrichment) were also co-assembled using identical parameters. Scaffolds ≥ 2 kb long were binned with MetaBAT2 v2.12 [41], Maxbin v2.2.6 [42] and CONCOCT v1.0.0 [43]. Best-scoring bins per assembly were selected with DAS_Tool v1.1.1 [44]. Bins were de-replicated across assemblies using dRep v2.0.5 (average nucleotide identity, ANI $> 99\%$; completeness $> 50\%$) [45], and manually refined using *t*-SNE transformation of tetranucleotide frequencies and coverage (https://github.com/dwwaite/bin_detangling). Genome completeness was estimated using CheckM [46]. For genome coverage, trimmed reads were mapped onto de-replicated genomes using bowtie2 v2.3.2 [47] (-n 1 -l 222 --minins 200 --maxins 800 -best). Sample-specific genome relative abundance was calculated by normalizing to library size and highest read count [48]. Estimated genome size was calculated as (bin size - (bin size * contamination))/(completeness) [49].

Metabolic predictions

Protein-coding gene sequences were predicted using Prodigal v2.6.3 [50] and annotated using USEARCH v9.02132 [36] with -usearch_global (-id 0.5 -evaluate 0.001 -maxhits 10), and UniRef100, UniProt [51] and KEGG databases [52]. Hidden Markov Model (HMM) searches were carried out using HMMER v3.3[53] against PFAM [54], TIGRfam [55] databases and databases from Anantharaman *et al.* [56] (HMM individual cutoffs, Table S4). BLASTP (NCBI, National Center for Biotechnology Information) was used to identify predicted protein-coding sequences for Rieske/cytochrome-b systems and the electron transfer module from *Candidatus* Kuenenia stuttgartiensis (similarity >30%, query coverage >70%) (Tables S4-5).

Genome classification and reconstruction/recovery of 16S rRNA gene sequences

Metagenome-assembled genomes (MAGs) were taxonomically classified using the Genome Taxonomy Database taxonomic classification tool, GTDB-Tk v0.2.1 [57]. 16S rRNA gene sequences were reconstructed from metagenomic data using EMIRGE [31] or SPAdes [40] with identification by Metaxa2 [58] and PATRIC [59]. EMIRGE was used to reconstruct 16S rRNA gene sequences from metagenomes with 97% cluster identity, 80 iterations and the SILVA SSU Ref NR99 database v132 [32].

Phylogenetic and protein sequence trees

Bacterial core gene alignments (75-114 genes per genome, Table S3) generated from GTDB-Tk were used to construct a maximum-likelihood phylogenomic tree in IQ-TREE (v1.6.9) [60] using ModelFinder [61] best-fit model LG+F+R5, and annotated with iTOL [62]. EMIRGE, Metaxa2 and amplicon sequences were aligned using MUSCLE [63] with default parameters and trimmed to 295 bp using Geneious 11.1.2 (<https://www.geneious.com>). A maximum-likelihood tree was constructed as with IQ-TREE using ModelFinder best-fit model TIM3+F+I+G4. Protein sequences of hydrazine synthase subunits A, B and C were aligned with MUSCLE and trimmed to remove columns with > 50% gaps using trimAl [64]. A maximum-likelihood tree was constructed as with IQ-TREE using ModelFinder best-fit model WAG+G4.

Metatranscriptome processing

Adapters were removed from transcriptomic reads, and quality trimmed and checked as described above for metagenomic reads. Residual ribosomal RNA sequences were removed using SortMeRNA v2.1 [65]. Additional checks were carried out to ensure filtered reads were still paired and read pairs were ordered using a repair script from BBMap v38.81 [66]. Filtered transcriptomic reads were mapped to contigs from the set of dereplicated MAGs using Bowtie2 [47] (v2.3.5, --end-to-end --very_sensitive). Read counts were determined using featureCounts [67] (v1.6.3, -F SAF). Singleton reads per gene were removed, and the remaining read counts were normalized to transcripts per kilobase per million reads mapped (TPM) using the following equation: $(\text{number of reads mapped to gene}) \times (1000 / \text{gene length}) \times (1000000 / \text{library size})$ [68].

Statistical analyses

Analyses were carried out in RStudio (v4.0.3) [69] with the packages: vegan v2.5.6 [70] (for distance-based redundancy analysis, dbRDA) and phyloseq v1.34.0 [71] (alpha diversity analysis). All correlations were Spearman's rank correlations (p -value <0.05 was considered significant).

Results and discussion

Prevalence and activity of anammox in chemically diverse groundwaters

Anammox bacteria are widespread in groundwater and phylogenetically diverse. All known anammox bacteria exclusively associate with order *Candidatus Brocadiales* [72] in the class *Candidatus Brocadiae* (*Brocadiae*). Based on 16S rRNA gene amplicon analysis of 80 groundwater samples collected across 10 aquifers, *Brocadiae* was present in 60 samples (from 47 wells) across all four regions assessed in this study. Results demonstrate *Brocadiae* prevalence within aquifers comprising unconsolidated materials (sandy-gravel=55/71, sand/silt=1/1), which are a common aquifer type globally [73], and consolidated volcanic rock (basalt/ignimbrite=4/7) (Fig. 1a, Table S1). Furthermore, *Brocadiae* was present in

groundwaters encompassing a range of nitrate-N (0–22 g/m³), DO (0–10.6 mg/L), phosphate (0–16.4 g/m³) and dissolved organic carbon (DOC, 0–26 mg/L) concentrations (based on bulk groundwater measurements, Table S1). The wide range of groundwater DO concentrations they were detected across is consistent with findings from two pristine carbonate-rock aquifers in Germany [8]. There, anammox was presumed to outcompete denitrification due to low organic carbon, and contributed to 83% of total N loss. Overall, *Brocadia* represented 0.37% of groundwater prokaryotic communities (3 940/1 071 440 sequences; ~49.3 sequences ± 202 SD per sample), and ranked 36th in abundance out of all 205 classes of bacteria and archaea. Research has shown that anammox is globally common in soils overlying aquifers, periodically saturated by high groundwater tables [16]. Results here demonstrate anammox-related bacteria are also common in the sediments and rock constituting aquifer saturated zones.

The class *Brocadia* comprised 37 OTUs (Table S7), of which 42% were assigned to four uncultivated or enrichment-cultivated genera (Fig. S1): *Candidatus Brocadia* (6 OTUs), *Candidatus Kuenenia* (1 OTU), *Planctomycetes* bacterium RIFCSPHIGH02_02_FULL_50_42 (6 OTUs), and *Planctomycetes* bacterium SM23_32 (3 OTUs). Over half (56.8%) could only be classified at the family (*Brocadia*ceae, 10 OTUs; GWA2-50-13, 9 OTUs) and order (*Brocadia*les, 2 OTUs) levels, highlighting a large degree of novelty among groundwater anammox candidates. Three of these taxonomic groups were strikingly abundant and prevalent, including the 7 most abundant OTUs overall, and representing >90% of all anammox sequences, *Ca. Brocadia*, RIFCSPHIGH02_02_FULL_50_42, and those classified only as *Brocadia*ceae (Fig. 1b). Together with previous studies, results show that the genus *Ca. Brocadia* is common in terrestrial subsurface environments [8], dominating groundwater anammox communities by up to 80% [74,75]. They were present in 83% of our samples, and comprised 25% of 16S rRNA gene sequences from anammox candidates. The wide distribution of *Ca. Brocadia* across varied groundwater conditions, as determined here, may be attributed to its growth properties (inferred from Monod models) [5]. *Ca. Brocadia* species exhibited the highest

maximum specific growth rate among anammox bacteria, and outcompete other freshwater anammox bacteria [5].

Brocadia relative abundance and community composition varies strongly with dissolved oxygen. The overall proportional abundance of *Brocadia* 16S rRNA gene sequences was significantly and negatively correlated with DO concentrations ($r=-0.28$) and ORP ($r=-0.27$; Table S8), a trend also demonstrated in batch culture experiments [76] and observed in carbonate-rock aquifers [8]. Significant positive correlations were instead found with DOC ($r=0.29$), TDS ($r=0.23$), conductivity ($r=0.27$), DRP ($r=0.31$) and NO_2^- ($r=0.31$, although only 14% of samples had concentrations above the detection limit for NO_2^-). TDS, conductivity, DRP and NO_2^- concentrations negatively correlated with DO overall ($p < 0.05$). Correlations between *Brocadia* and other substrates, including those critical for ammonia oxidation (NH_4^+ and NO_3^-), were not significant, indicating that parameters such as DO, ORP, nitrite and DOC are more important factors for controlling the presence of anammox bacteria. Of these parameters, multivariate regression tree analysis, showed that redox potential (ORP threshold 175.9 mV) was the main factor discriminating *Brocadia* communities (Fig. S2).

Analysis of anammox community diversity by oxygenation regime (anoxic, suboxic, dysoxic and oxic) revealed OTU richness and alpha-diversity (Inverse-Simpson) were significantly higher at the suboxic and dysoxic sites compared with oxic sites (Fig. 2a), suggesting groundwater with low oxygen supports more anammox species. The spatial distribution of different *Brocadia* taxa was significantly associated with changes in DO and bore depth (db-RDA, Fig. 2b), along with aquifer location and lithology, phosphate, and pH, which collectively explained 16% (R^2 adjusted) of the variation (Permutation-test, $p < 0.05$, permutations=999, Table 8). When considering only DO, dramatic differences in composition were evident (Fig. 1b). Notably, the *Ca. Brocadia* fraction was positively correlated with DO concentrations ($r=0.46$) (Fig. 1b). Three *Ca. Brocadia* and *Brocadia* OTUs comprised up to 30.4—89% of the anammox-associated community in groundwater with >4 mg/L DO (Figs S3-4).

Oxygen tolerance in anammox bacteria has been demonstrated previously [77,78], particularly for *Ca. Brocadia*. For example, *Ca. Brocadia sinica* exhibited tolerance in the presence of >1 mg/L DO *in vitro* [79], and anammox bacteria closely related to *Ca. Brocadia fulgida* have been reported in groundwater with up to 3 mg/L oxygen [8] (>3 mg/L DO is generally classified as oxic [29]). A recent study analysing the metabolism of *Ca. Brocadia* species, identified genes encoding superoxide dismutase and cytochrome c peroxidase genes that are involved in oxygen detoxification [80], as may be expected for organisms living at reduction-oxidation interfaces. Although *Brocadia* overall and *Ca. Brocadia* were detected in groundwater with much higher DO concentrations in this study (up to 10.6 mg/L), aquifers are spatially heterogeneous, and *Brocadia* likely inhabit lower oxygen niches within the sampled aquifers. In comparison, a further 25.9% of *Brocadia* comprised *Planctomycetes* RIFCSPHIGH02_02_FULL_50_42 (mostly OTU 75). The overall relative abundance was significantly and negatively correlated with DO ($r=-0.38$), ORP ($r=-0.45$) and NO_3^- ($r=-0.29$), indicating a preference for reducing conditions (Table S8). It was originally recovered from an aquifer adjacent to the Colorado River (Rifle, USA) [56], with conditions similar to site (Wel13) here, where the genus was most abundant, including identical dysoxic conditions (0.78 mg/L DO) with low NO_2^- and NO_3^- .

DOC did not negatively impact Brocadia community abundance. Anammox generally occurs at low organic carbon levels [81]. As reported above, DOC and *Brocadia* community abundance were instead positively correlated. This association may be best explained by the negative relationship typically shared by DO and DOC in groundwater, where organic carbon availability regulates aerobic metabolism [82], although DO and DOC were not significantly correlated in this study ($r=-0.08$, $p=0.5$), possibly due to DOC being rapidly utilized in groundwater. Nevertheless, some anammox bacteria (e.g. *Ca. Anammoxoglobus propionicus*) have a better affinity for NO_2^- in the presence of organic acids, such as propionate, or can assimilate formate via the Wood–Ljungdahl pathway (as recently shown for *Ca. Kuenenia*

stuttgartiensis) [83], implying that groundwater DOC could support the growth of some anammox bacteria.

Hydrazine synthase capacity, but not activity, was correlated with DO. To confirm anammox potential and activity in groundwater, we quantified genes and transcripts encoding hydrazine synthase subunit *hzsB*. Together, genes *hzsA*, *hzsB* and *hzsC* encode hydrazine synthase, which is a key enzyme in anammox metabolism, converting nitric oxide and ammonium into hydrazine [6]. The relative abundance of *Brocadia* 16S rRNA gene sequences, correlated strongly with *hzsB* gene copies ($r=0.62$) (Fig. 3a), consistent with the expectation that *Brocadia* undertake anammox [84]. The concentration of *hzsB* genes ranged from 3.64×10^1 to 1.23×10^7 genes/L of groundwater (excluding 11.1% of samples below detection) with an average of $4.9 \times 10^5 \pm 2.1 \times 10^6$ SD genes/L. This is consistent with concentrations of $\sim 1.5 \times 10^4$ to $\sim 1.0 \times 10^8$ genes/L found elsewhere in the subsurface, including in soil horizons interacting with the water table [8,85]. The average number of *hzsB* transcript copies was significantly lower than for *hzsB* gene copies (Wilcoxon test $p < 0.01$; 4.5×10^5 copies/L less, ~ 13.2 times lower) (Fig. 3b), suggesting a latent capacity of 92.4%. Overall, *hzsB* gene copy numbers were negatively correlated with bore depth ($r=-0.34$) and DO ($r=-0.42$), reflecting the trend seen with *Brocadia* 16S rRNA gene relative abundance and DO, and as expected for an anaerobic process. However, no significant relationship was evident between *hzsB* gene transcripts or anammox activity and oxygen availability ($r=-0.16$, $p=0.44$) or bore depth ($r=0.14$, $p=0.5$).

Co-occurrence of aerobic and anaerobic ammonia oxidizers

The nitrification-anammox process allows for elevated N-removal efficiencies from NH_4^+ in naturally ammonium-rich or contaminated water [86], by ensuring a portion of ammonium is converted to nitrite. Potential combined nitrification-anammox was determined by quantification of archaeal and bacterial ammonia monooxygenase genes (*amoA*). Results show *hzsB* and archaeal and bacterial *amoA* genes ($r=0.47$) and transcripts ($r=0.72$) were strongly correlated, suggesting co-occurrence of aerobic and anaerobic ammonium oxidation

(Fig. 3c-d). Co-occurrence has been observed previously within lab-based models [87], and an aquifer system, where anammox was the dominant process [8]. Together, these results suggest that anammox in aquifers relies, at least partially, on nitrite derived from aerobic ammonia oxidation. Results here also indicate anammox is more frequently the dominant process when considering a wide range of aquifers. Aerobic ammonia monooxygenase gene copies were more abundant in 83% of samples compared to hydrazine synthase genes; however, hydrazine synthase transcripts were higher in 69% of samples.

Phylogenomic diversity of anammox bacteria in aquifers

Genomes were reconstructed from 16 samples (gwj01-gwj16) across 8 wells (groundwater $\pm$ biomass-enrichment) spanning a range of oxygen (0.37–7.5 mg/L) and nitrate (0.47–12.6 g/m³) concentrations (Table S1). Of 541 MAGs (>50% complete, $\leq$ 5% contamination), 8 were identified as *Ca. Brocadia* (63.4–95.6% complete, Table S10). Of these MAGs that contained partial or (near)full length (445-1496 bp) 16S rRNA gene sequences (nzgw513, nzgw516-517), those from nzgw516-517 aligned fully with OTU729 (272 bp long), which was present in 21/80 samples (total of 92 sequences). Both the nzgw516-517 and OTU729 sequences had best hits to *Planctomycetes* spp., including *Planctomycetes* bacterium strain Pla86 and *Ca. Kuenenia stuttgartiensis* when searched against the NCBI NR database (BLASTN; 79.54%-82.85% identity), consistent with phylogenomic analyses showing these genomes comprise a novel clade (Clade II, Fig. 4a). OTU729 was only classified to the domain Bacteria using USEARCH with the SILVA SSU Ref NR99 database, and RDP Classifier, suggesting that these novel *Ca. Brocadia* could be underestimated in the environment when using a 16S rRNA gene based approach.

A pairwise comparison with 16 other *Brocadia* genomes from various environments yielded ANI values below the proposed species cut off of 95% [88] (Fig. S5), implying the newly recovered MAGs represent novel species. The phylogenomic tree (Fig. 4) revealed six NZ aquifer-derived MAGs form three (sub-)clades with recently recovered genomes from the aquifer in Rifle, Colorado [56], suggesting these are groundwater adapted species. One

(nzwg512) clustered with well-characterised anammox bacterium, *Ca. Jettenia caeni* [89]. An eighth MAG, nzwg510, was phylogenetically distinct from other *Ca. Brocadia* MAGs clustering with Rifle MAG “MHYJ01” (from aquifer sediment [56]) constitute Clade-II, which is distinct from well-known anammox bacteria (Clade-I, Fig. 4a). They contained a much higher GC content (GC >67%) than others from this study, excluding outlier nzwg510 (average 42.5% $\pm$ 3 SD) (Table S10). Higher GC may indicate a higher evolutionary rate for these organisms [90]. All lower GC *Brocadia* MAGs from this study, including two aquifer subclades, co-clustered with well-known anammox *Candidatus* genera from enrichment studies (Clade-I). The estimated genome sizes were significantly smaller than the high GC group (Clade-I: 3.22 Mbp $\pm$ 0.92, Clade-II: 5.16 Mbp $\pm$ 1.24), further suggesting a divergence in evolutionary strategy employed by the two groups.

As for 16S rRNA gene analyses (Fig. 1b), *Brocadia* MAG relative abundances varied across sites, corresponding to distinct geochemical conditions (Fig.4a, Table S10) but not strictly phylogenetic relatedness (Fig. 4b). Three of four MAGs associated with well-characterized Clade I anammox-bacteria (nzwg511-513) were most abundant at sites with low DO (<1.1 mg/L) and high DOC (3–26 g/m³). MAG nzwg514 is one of two closely related to *Ca. Brocadia*. Consistent with 16S rRNA gene data (Fig. 1b) it demonstrated a preference for oxic groundwater (Fig. 4b). Together with the high-GC MAGs (nzwg510 and Clade II), it was more abundant at sites with high DO (5.6–7.5 mg/L) and low DOC (0–0.8 g/m³). Factors identified that influence ecological niche differentiation among anammox species, include microbial growth kinetics, organic matter, oxygen tolerance, aggregation ability and interspecific competition [5]. Collectively, results here suggest DO availability may be a major driver of anammox bacterial diversification in groundwater.

Genomic and transcriptomic evidence for anammox metabolism and oxygen tolerance

Characteristic hydrazine based steps in anammox. Nitric oxide produced by NO₂⁻ reductase [91], or by cyclic feeding, is used by hydrazine synthase (HZS), along with NH₄⁺, to produce hydrazine (N₂H₄) within the anammoxosome [6]. Most *Brocadia* MAGs recovered (6 of 8)

have genes encoding signature anammox steps for hydrazine production (hydrazine synthase) and removal (hydrazine dehydrogenase) (Fig. 6; Table S11) [91]. This excludes phylogenetic outlier nzgw510 and Clade-II nzgw517, with the lowest genome completeness (75% and 63%, respectively). Nonetheless, HZS genes in nzgw515-516 indicate that Clade-II bacteria are capable of anammox [91], and may be important contributors to anammox in oxygen-rich groundwaters (Fig. 4b) [8]. The final anammox step is catalysed by an octahaem-hydrazine dehydrogenase (HDH/*hzoA*). N_2H_4 is oxidised using cytochrome *c* as an electron acceptor [92]. Although no Clade-II MAGs contained *hzoA*, nzgw515 contained hydroxylamine oxidoreductase, which is also capable of hydrazine oxidation *in vitro* [93]. Multiple *hzoA* gene copies are present in nzgw511–514, consistent with close phylogenetic relatedness to well-characterised anammox bacteria (Fig. 4).

Reinforcing genomic evidence, metatranscriptomics data for six samples (gwj09, gwj11, gwj13-16) across two sites, revealed that the transcriptional activity of *hzoA* in MAGs nzgw511-514. This activity was 356-fold higher at the dysoxic site (wells E1 and N3) than the oxic site (Fig. 6b), and likely contributed in part to excess N_2 measured at those sites (Fig. 5). Despite the compositional bias in anammox bacteria associated with differences in DO (Fig. 4), results confirm that in groundwater with low oxygen concentrations anammox bacteria are more active [8]. Contemporaneous measurement of excess N_2 indicated active N_2 generation in dysoxic groundwater (from wells E1 and N3) due to denitrification or anammox. In contrast, oxic groundwater was either at the bounds of uncertainty (wells BW8 and RF3) or devoid of measurable excess N_2 (from wells SR1-2, BW19, RF2) (Fig. 5), which, accordingly, included groundwater with relatively little observed hydrazine synthase/dehydrogenase activity (from wells SR1-2). Additional quantification via ddPCR of hydrazine synthase transcripts from each site, sampled at an earlier date, demonstrated that anammox was active in groundwater associated with all *Ca. Brocadia* MAGs, which were present in both dysoxic to oxic sites (Fig. 4b) distributed across 8 wells over 4 sites (Table S9). However, quantitative results were

consistent with relative metatranscriptomics data, indicating transcriptional activity was 100-fold higher in dysoxic versus oxic groundwater in the four samples analyzed (Fig. S6).

Further analyses showed HZS sequence similarity and structure did not reflect *Brocadia* phylogenomic relatedness. All three HZS subunits encoded by nzgw511 are closely related to *Ca. Kuenenia stuttgartiensis* subunits (Fig. 7). However, HzsB from Clade-II nzgw516, and Clade-I nzgw514, and nzgw511 (duplicate), constitute a distinct HzsB clade relative to other anammox bacteria, indicating evolutionary divergence. Clade-II nzgw515 instead encodes for a fused HZS-beta-gamma protein subunit, comparable to distant *Brocadia* relatives *Ca. Scalindua profunda* [94] and *Scalindua brodae* [95].

Production of nitric oxide (and ammonium) for anammox. In the typical anammox reaction, the first step is NO_2^- reduction to NO. Nitrite, in addition to other external substrates, such as nitrate, nitrite and ammonium, are acquired by transporters (see Supplementary Information for details on transporters in MAGs). NO_2^- reduction to NO is usually catalysed by cytochrome *cd*₁-type nitrite reductase (*nirS*), as found in *Ca. Kuenenia* and *Ca. Scalindua* [94,96], or the copper-containing nitrite reductase (*nirK*), as for *Ca. Jettenia* spp. [97]. One MAG (nzgw512) contained *nirK* similar to *Ca. Jettenia caeni* [89]; however, other MAGs were devoid of *nirK* and *nirS* (Fig. 6). Canonical NO-forming NO_2^- reductases are also lacking in several *Brocadia* species (e.g. *Ca. Brocadia sapporoensis* and *Ca. Brocadia sinica*) [98] and Rifle aquifer genomes (e.g. 2-12-FULL-39-13 and 2-02-FULL-50-16-A) [56]. Some anammox bacteria are proposed to utilize different enzymes to reduce NO_2^- to NO [99]. These enzymes may be among the hydroxylamine oxidoreductase like octaheme proteins encoded by anammox genomes [91]. Within *Ca. Kuenenia stuttgartiensis*, this protein (kust0457-kust0458), comprising octaheme kustc0458 and diheme kustc0457, constitutes a heterododecameric ($\alpha_6\beta_6$) complex comprising 60 c-type hemes [91]. The second is kuste4574, a close homolog of kustc0458, which forms part of a novel Rieske-heme *b* (R/b; *bc*₁) complex (kuste4569-kuste4574). It is postulated that these proteins catalyse NO_2^- reduction to NO [91]. Protein-coding sequences similar to kustc0457-kustc0458

and kuste4574 were identified in nzgw511, nzgw513-514 (Table S6). These proteins may represent an alternative pathway for reducing NO_2^- . Furthermore, *Ca. Kuenenia stuttgartiensis* can grow in the absence of NO_2^- by coupling ammonium oxidation to NO reduction, producing N_2 without N_2O emissions [100]. This suggests that anammox bacteria may mitigate natural NO and N_2O emissions in many ecosystems.

It is suggested that anammox bacteria adapt a Dissimilatory Nitrate Reduction to Ammonia (DNRA) like metabolism in the absence of NH_4^+ , using nitrate reductase, followed by NrfAH cytochrome *c* nitrite reductase [72]. The respective products, NO_2^- and NH_4^+ , can then drive anammox. We found *nrfAH* genes in phylogenetically diverse genomes (nzgw514, nzgw516-517). Further, genomes nzgw512–514 contain a tetraheme cytochrome *c nrfH* gene, with CxxCH repeated sequence motifs similar to *Ca. Brocadia* [99] and *Ca. Jettenia* [97]. Both *nrfA* and *nrfH* genes were observed to be expressed by these taxa in dysoxic groundwater (Fig. 6b). Together, these genes and transcripts indicate the active reduction of NO_2^- to NH_4^+ in a NAD(P)H-dependent, six-electron transfer reaction, via a mechanism resembling DNRA [91], potentially fuelling active anammox in the same sites.

NAD(P)H nitrite reductase genes (*nirD* and/or *nirB*), which also reduce NO_2^- to NH_4^+ , were found in Clade-II genomes, nzgw515–517. The large subunit encodes NirB, an iron–sulfur protein. All protein-coding sequences had the highest identity (79–84%) to NirB from *Planctomycetes* genome (GCA_016200125.1), recently recovered from a pristine aquifer in the USA [101]. This had a similar geochemical profile to the site where nzgw516-517 were most abundant (low NH_4^+ , <0.05 mg/L, and NO_3^- , 0.23 mg/L). The *nirB* gene was recently also identified in a marine-derived anammox bacterium, related to *Ca. Scalindua* [102]. We also identified a small subunit, NirD, in nzgw516-517. The sequence contains a Rieske non-heme iron oxygenase family domain with closest similarity to *Verrucomicrobia* species (63.5%). NH_4^+ often limiting in groundwater [20]. In this study, NH_4^+ was undetectable in groundwater from which MAGs were recovered, and detected at only 31% of sites overall. The presence of these

Nrf and Nir proteins could therefore mediate an important alternative process providing NH_4^+ for anammox [91].

Mechanisms of oxygen tolerance in groundwater anammox bacteria. All *Brocadiae* genomes encode proteins for oxygen protection, such as BatD and BatA, which are involved in oxygen tolerance in *Bacteroides* species [103], alkyl hydroperoxide reductases and thioredoxin [104]. Thioredoxin and peroxiredoxin genes were expressed in MAG nzgw510 at the oxic site (0.06-0.15 TPM), and BatD, BatA, catalase, peroxiredoxin and thioredoxin were expressed by MAGs nzgw511-514 at the dysoxic site (0.112-2.81 TPM). Proteins encoding respiratory chain cytochrome *c* oxidase, *cbb3*-type (*ccoN*, *ccoO* and *ccoP*) were all or partially present in four MAGs (nzgw511-14). *Cbb3* is a multi-chain transmembrane protein located near the anammoxosome membrane [105]. This enzyme is believed to have evolved to perform a specialized function in microaerobic energy metabolism, however it could also merely act in oxygen detoxification [72]. Additionally, three Clade-II genomes encode a *caa3*-type cytochrome oxidase, containing unique *CoxA/B* subunits. In *Acidithiobacillus ferrooxidans* *coxA/B* genes encode the terminal part of the ferrous iron “downhill” pathway, which shuttles electrons from extracellular iron to oxygen [106], and thus represents another potential detoxification strategy.

Alternative metabolic pathways in groundwater anammox bacteria. Anammox grow chemolithoautotrophically by performing CO_2 fixation via the Wood-Ljungdahl pathway [107]. The products of this reaction, acetyl-CoA and pyruvate, enter the oxidative TCA cycle and gluconeogenesis (Fig. 6; Table S10; see details on TCA cycle in Supplementary Information). Several ABC-transport systems were present within the genomes, revealing variations in substrate importation such as phosphate, cobalt, nickel, iron (III), zinc, sulfate, molybdate, lipoproteins, ribose, rhamnose, polysaccharides and oligopeptides (Table S10). Oligopeptide transport systems were present in three genomes (nzgw515-517), as seen in *Ca. Scalindua profunda*, and suggest these genomes are capable of oxidizing decaying organic matter [108], in addition to carbon fixation via the Wood-Ljungdahl pathway (Fig. 6). Additionally, iron (III)

transporters were present in three genomes which may enable the coupling of formate oxidation to iron (III) reduction in the absence of NO_2^- or NO , as previously identified in other species [109]. Moreover, hydrogenase genes were also present in three of the genomes. The first encodes a [Ni-Fe] group 3b hydrogenase, found in nzgw511, which is proposed to harbour (sulf)hydrogenase activity [110]. This genome also encoded a sulfate adenylytransferase, which may play a role in sulfide oxidation [111]. The second is a group 4 hydrogenase present in nzgw513. These hydrogenases encode coupled oxidation of NADPH to the evolution of H_2 [112], indicating hydrogen turnover and/or an alternative energy source. Our results therefore suggest these groundwater anammox bacteria are metabolically versatile, containing various hydrogenases and ABC transporters for organic compounds, and genes encoding DNRA, which would allow for growth in substrate limiting conditions (such as low ammonium concentrations).

Conclusions

This study shows that anammox bacteria are prevalent and active across wide-ranging aquifer chemistries and lithologies, including in oxic groundwater, although predicted to be unfavourable. While we found significantly more *Brocadiae* diversity in anoxic groundwater, some taxa were positively associated with DO concentrations. Of the eight novel *Brocadiae* genomes reconstructed, three belong to under-characterised clades only previously recovered from aquifers. Most possess genes involved in signature pathways of anammox, including novel hydrazine synthase genes. A co-occurrence of anaerobic and aerobic ammonia-oxidisers at many sites suggests metabolic hand offs (such as nitrite) between these processes. Furthermore, genomic characterization of *Brocadiae* identified a range of potential aero-tolerance mechanisms, explaining our finding of anammox in oxic groundwaters. Results indicate niche differentiation amongst anammox bacteria based on oxygen concentrations, and that anammox is a common mechanism for nitrogen removal in aquifers.

Declarations

578 **Ethics approval and consent to participate**

579 Not applicable.

580 **Consent for publication**

581 Not applicable.

582 **Availability of data and materials**

583 All amplicon, WGS and metatranscriptome data from groundwater and biomass-enriched
584 groundwater samples have been deposited with NCBI under BioProject PRJNA699054.

585 **Competing interests**

586 The authors declare no conflict of interest.

587 **Funding**

588 Research was supported by a Smart Ideas grant, from the Ministry of Business, Innovation
589 and Employment, awarded to KMH (project UOAX1720).

590 **Authors' information**

591 Affiliations

592 **School of Biological Sciences, The University of Auckland, Auckland, New Zealand**

593 Olivia E. Mosley, Emilie Gios & Kim M. Handley

594 **Institute of Environmental Science and Research, Christchurch, New Zealand**

595 Louise Weaver & Murray Close

596 **National Institute of Water and Atmospheric Research, Wellington, New Zealand**

597 Chris Daughney

598 **GNS Sciences, Lower Hutt, New Zealand**

599 Rob van der Raaij & Heather Martindale

600 **Authors' contributions**

OM, EG and KMH collected samples. OM and EG led field trips. OM, EG, KMH, LW and MC planned the study. OM and EG conducted the laboratory work. OM and EG analyzed and interpreted data. RR and HM quantified excess N₂ gas at four sites. OM wrote the first draft of the manuscript. OM, KMH and EG revised and corrected the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Kim M. Handley.

Acknowledgements

We thank P Abraham (ESR), HS Tee and JS Boey (University of Auckland) for help sampling Canterbury wells, and council staff in Auckland (C Foster), Wellington (D McQueen), Waikato (S Herath) and Canterbury (D Evans and R Cressy) for help accessing and sampling wells. We thank D Waite and JS Boey for bioinformatics support. Computational resources were provided by New Zealand eScience Infrastructure.

References

1. Rivett MO, Buss SR, Morgan P, Smith JWN, Bemment CD. Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. *Water Res.* 2008;42(16):4215–32. <https://doi.org/10.1016/j.watres.2008.07.020>.
2. Jetten MSM. The microbial nitrogen cycle. *Environ. Microbiol.* 2008;10(11):2903–9. <https://doi.org/10.1111/j.1462-2920.2008.01786.x>.
3. Strous M, Fuerst JA, Kramer EHM, Logemann S, Muyzer G, Van De Pas-Schoonen KT, et al. Missing lithotroph identified as new planctomycete. *Nature.* 1999;400:446–9. <https://doi.org/10.1038/22749>.
4. Oshiki M, Satoh H, Okabe S. Ecology and physiology of anaerobic ammonium oxidizing bacteria. *Environ Microbiol.* 2016;18(9):2784–96. <https://doi.org/10.1111/1462-2920.13134>.

5. Zhang L, Okabe S. Ecological niche differentiation among anammox bacteria. *Water Res.* 2020;171:115468. <https://doi.org/10.1016/j.watres.2020.115468>.
6. Kartal B, Maalcke WJ, De Almeida NM, Cirpus I, Gloerich J, Geerts W, et al. Molecular mechanism of anaerobic ammonium oxidation. *Nature.* 2011;479(7371):127–30. <https://doi.org/10.1038/nature10453>
7. van Niftrik L, Jetten MSM. Anaerobic ammonium-oxidizing bacteria: Unique microorganisms with exceptional properties. *Microbiol Mol Biol Rev.* 2012;76(3):585–96. <https://doi.org/10.1128/MMBR.05025-11>.
8. Kumar S, Herrmann M, Thamdrup B, Schwab VF, Geesink P, Trumbore SE, et al. Nitrogen loss from pristine carbonate-rock aquifers of the hainich critical zone exploratory (Germany) is primarily driven by chemolithoautotrophic anammox processes. *Front Microbiol.* 2017;8:1951. <https://doi.org/10.3389/fmicb.2017.01951>.
9. Kartal B, Rattray J, van Niftrik LA, van de Vossenberg J, Schmid MC, Webb RI, et al. Candidatus “Anammoxoglobus propionicus” a new propionate oxidizing species of anaerobic ammonium oxidizing bacteria; 2007;30(1):39–49. <https://doi.org/10.1016/j.syapm.2006.03.004>.
10. Dalsgaard T, Thamdrup B, Farías L, Revsbech NP. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol Oceanogr.* 2012;57(5):1331–46. <https://doi.org/10.4319/lo.2012.57.5.1331>.
11. Kalvelage T, Jensen MM, Contreras S, Revsbech NP, Lam P, Günter M, et al. Oxygen sensitivity of anammox and coupled N-cycle processes in oxygen minimum zones. *PLoS One.* 2011;6(12):e29299. <https://doi.org/10.1371/journal.pone.0029299>.
12. Schubert CJ, Durisch-Kaiser E, Wehrli B, Thamdrup B, Lam P, Kuypers MMM. Anaerobic ammonium oxidation in a tropical freshwater system (Lake Tanganyika). *Environ Microbiol.* 2006;8(10):1857–63. <https://doi.org/10.1111/j.1462-2920.2006.01074.x>.

13. Rich JJ, Dale OR, Song B, Ward BB. Anaerobic ammonium oxidation (anammox) in Chesapeake Bay sediments. *Microb Ecol.* 2008;55(2):311–20.
<https://doi.org/10.1007/s00248-007-9277-3>.
14. Hanson BT, Madsen EL. In situ expression of nitrite-dependent anaerobic methane oxidation proteins by *Candidatus Methyloirabilis oxyfera* co-occurring with expressed anammox proteins in a contaminated aquifer. *Environ Microbiol Rep.* 2015;7(12):252–64.
<https://doi.org/10.1111/1758-2229.12239>.
15. Smith RL, Böhlke JK, Song B, Tobias CR. Role of anaerobic ammonium oxidation (anammox) in nitrogen removal from a freshwater aquifer. *Environ Sci Technol.* 2015;49(20):12169–77. <https://doi.org/10.1021/acs.est.5b02488>.
16. Wang S, Zhu G, Zhuang L, Li Y, Liu L, Lavik G, et al. Anaerobic ammonium oxidation is a major N-sink in aquifer systems around the world. *ISME J.* 2019;14:151–163.
<https://doi.org/10.1038/s41396-019-0513-x>.
17. Aeschbach-Hertig W, Gleeson T. Regional strategies for the accelerating global problem of groundwater depletion. *Nat Geosci.* 2012;5:853–61.
<https://doi.org/10.1038/ngeo1617>.
18. Strous M, Kuenen JG, Jetten MSM. Key physiology of anaerobic ammonium oxidation. *Appl Environ Microbiol.* 1999;65(7):3248–50. <https://doi.org/10.1128/AEM.65.7.3248-3250.1999>.
19. Morgenstern U, Daughney CJ. Groundwater age for identification of baseline groundwater quality and impacts of land-use intensification - The National Groundwater Monitoring Programme of New Zealand. *J Hydrol.* 2012;456–457:79–93.
20. Nitrate and nitrite in drinking-water Background document for development of WHO Guidelines for Drinking-water Quality. *World Health Organization* 2003:
<https://apps.who.int/iris/handle/10665/75380>.

21. Hansen B, Thorling L, Schullehner J, Termansen M, Dalgaard T. Groundwater nitrate response to sustainable nitrogen management. *Sci Rep.* 2017;7:8566.
<https://doi.org/10.1038/s41598-017-07147-2>.
22. Leenheer JA, Malcolm RL, McKinley PW, Eccles LA. Occurrence of dissolved organic carbon in selected ground-water samples in the United States. *J Res US Geol Surv.* 1974;2.
23. Zhu G, Wang S, Wang W, Wang Y, Zhou L, Jiang B, et al. Hotspots of anaerobic ammonium oxidation at land-freshwater interfaces. *Nat Geosci.* 2013;6:103–7.
<https://doi.org/10.1038/ngeo1683>.
24. Kuenen JG. Anammox and beyond. *Environ Microbiol.* 2020;22(2):525–36.
<https://doi.org/10.1111/1462-2920.14904>.
25. Schmidt CM, Fisher AT, Racz AJ, Lockwood BS, Huertos ML. Linking denitrification and infiltration rates during managed groundwater recharge. *Environ Sci Technol.* 2011;45(22):9634–40. <https://doi.org/10.1021/es2023626>.
26. Nieder R, Benbi DK, Scherer HW. Fixation and defixation of ammonium in soils: A review. *Biol Fertil Soils* 2011;47:1–14. <https://doi.org/10.1007/s00374-010-0506-4>.
27. Monteiro M, Sneca J, Magalhes C. The history of aerobic ammonia oxidizers: From the first discoveries to today. *J. Microbiol.* 2014;52(7):537–47.
<https://doi.org/10.1007/s12275-014-4114-0>.
28. Close M, Abraham P, Webber J, Cowey E, Humphries B, Fenwick G, et al. Use of sonication for enhanced sampling of attached microbes from groundwater systems. *Groundwater*; 2020;58:901–12. <https://doi.org/10.1111/gwat.12984>.
29. Malard F, Hervant F. Oxygen supply and the adaptations of animals in groundwater. *Freshw Biol.* 1999;41(1):1–30. <https://doi.org/10.1046/j.1365-2427.1999.00379.x>.
30. Martindale H, van der Raaij R, Daughney CJ, Morgenstern U, Singh R, Jha N, et al. Assessment of excess N₂ for quantifying actual denitrification in New Zealand groundwater systems. *J Hydrol New Zeal.* 2019;58(1):1–17.

31. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: Reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 2011;12:R44.
32. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. 2013;41:D590-6. <https://doi.org/10.1093/nar/gks1219>.
33. Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S, et al. DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl Environ Microbiol.* 2014;80(16):5116–23. <https://doi.org/10.1128/AEM.01403-14>.
34. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol.* 2015;75(2):129–37. <https://doi.org/10.3354/ame01753>.
35. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge. 2010. Available at <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
36. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460–1. <https://doi.org/10.1093/bioinformatics/btq461>.
37. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10(10):996–8. <https://doi.org/10.1038/nmeth.2604>.
38. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335–6. <https://doi.org/10.1038/nmeth.f.303>.
39. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>.

40. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
41. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359. <https://doi.org/10.7717/peerj.7359>
42. Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome.* 2014;2:26. <https://doi.org/10.1186/2049-2618-2-26>.
43. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*; 2014;11(11):1144–6. <https://doi.org/10.1038/nmeth.3103>.
44. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol*; 2018;3(7):836–43. <https://doi.org/10.1038/s41564-018-0171-1>.
45. Olm MR, Brown CT, Brooks B, Banfield JF. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 2017;11:2864–8. <https://doi.org/10.1038/ismej.2017.126>.
46. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25(7):1043–55. <https://doi.org/10.1101/gr.186072.114>.
47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
48. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol.* 2018;3(3):328–36. <https://doi.org/10.1038/s41564-017-0098-y>.

49. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat Rev Microbiol.* 2018;16(10):629–45. <https://doi.org/10.1038/s41579-018-0076-2>.
50. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:119. <https://doi.org/10.1186/1471-2105-11-119>.
51. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 2017;45:D158–69. <https://doi.org/10.1093/nar/gky1049>.
52. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 1999. <https://doi.org/10.1093/nar/27.1.29>.
53. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46(W1):W200–4. <https://doi.org/10.1093/nar/gky448>.
54. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. *Nucleic Acids Res.* 2014;42:D222-D230 <https://doi.org/10.1093/nar/gkt1223>.
55. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 2003;31(1):371–3. <https://doi.org/10.1093/nar/gkg128>.
56. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun.* 2016;7:1–11. <http://doi.org/10.1038/ncomms13219>.
57. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019;36(6):1925–1927. <http://doi.org/10.1093/bioinformatics/btz848>.

58. Bengtsson-Palme J, Thorell K, Wurzbacher C, Sjöling Å, Nilsson RH. Metaxa2 Diversity Tools: Easing microbial community analysis with Metaxa2. *Ecol Inform.* 2016;33:45–50. <http://doi.org/10.1016/j.ecoinf.2016.04.004>.
59. Brettin, T., Davis, J., Disz, T. et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep.* 2015;5:8365. <https://doi.org/10.1038/srep08365>
60. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>.
61. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285>.
62. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 2019;47(W1):W256-259. <https://doi.org/10.1093/nar/gkz239>.
63. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:113. <https://doi.org/10.1186/1471-2105-5-113>.
64. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
65. Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics.* 2012;28:3211–7. <https://doi.org/10.1093/bioinformatics/bts611>.
66. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory. LBNL-7065E. Retrieved from <https://escholarship.org/uc/item/1h3515gn>

67. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
<https://doi.org/10.1093/bioinformatics/btt656>.
68. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. 2012;131:281–5.
<https://doi.org/10.1007/s12064-012-0162-3>.
69. RStudio Team. RStudio server: Integrated Development for R. RStudio, Inc, Boston, MA. 2016;42:14.
70. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. *vegan: Community Ecology. R Packag*. 2019. Available from: <https://cran.r-project.org/web/packages/vegan/vegan.pdf>.
71. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. Watson M, editor. *PLoS One*. 2013;8:e61217.
<https://doi.org/10.1371/journal.pone.0061217>.
72. Kartal B, De Almeida NM, Maalcke WJ, Op den Camp HJM, Jetten MSM, Keltjens JT. How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol Rev*. 2013;37(3):428–61. <https://doi.org/10.1111/1574-6976.12014>.
73. Gleeson T, Befus KM, Jasechko S, Luijendijk E, Cardenas MB. The global volume and distribution of modern groundwater. *Nat Geosci*. 2016;9:161–4.
<https://doi.org/10.1038/ngeo2590>.
74. Sonthiphand P, Hall MW, Neufeld JD. Biogeography of anaerobic ammonia-oxidizing (anammox) bacteria. *Front Microbiol. Frontiers Research Foundation*; 2014;5:399.
<https://doi.org/10.3389/fmicb.2014.00399>.
75. Wang Y, Xu L, Wang S, Ye F, Zhu G. Global distribution of anaerobic ammonia oxidation (anammox) bacteria – field surveys in wetland, dryland, groundwater aquifer and snow. *Front Microbiol*. 2019;10:2583. <https://doi.org/10.3389/fmicb.2019.02583>.

76. Nguyen Viet T, Kumar Behera S, Kim JW, Park H-S. Effects of oxidation reduction potential and organic compounds on anammox reaction in batch cultures. *Environ Eng Res.* 2008;13(4)210–215.
77. Wang S, Peng Y, Ma B, Wang S, Zhu G. Anaerobic ammonium oxidation in traditional municipal wastewater treatment plants with low-strength ammonium loading: Widespread but overlooked. *Water Res.* 2015;84:66–75. <https://doi.org/10.1016/j.watres.2015.07.005>.
78. Seuntjens D, Carvajal-Arroyo JM, Ruopp M, Bunse P, De Mulder CP, Lochmatter S, et al. High-resolution mapping and modeling of anammox recovery from recurrent oxygen exposure. *Water Res.* 2018;144:522–31. <https://doi.org/10.1016/j.watres.2018.07.024>.
79. Oshiki M, Shimokawa M, Fujii N, Satoh H, Okabe S. Physiological characteristics of the anaerobic ammonium-oxidizing bacterium “*Candidatus Brocadia sinica*.” *Microbiology.* 2011;157(6):1706–13. <https://doi.org/10.1099/mic.0.048595-0>
80. Ji X, Wu Z, Sung S, Lee PH. Metagenomics and metatranscriptomics analyses reveal oxygen detoxification and mixotrophic potentials of an enriched anammox culture in a continuous stirred-tank reactor. *Water Res.* 2019;166:115039. <https://doi.org/10.1016/j.watres.2019.115039>.
81. Wu G, Zhang T, Gu M, Chen Z, Yin Q. Review of characteristics of anammox bacteria and strategies for anammox start-up for sustainable wastewater resource management. *Water Sci Technol.* 2020;82(9):1742–57. <https://doi.org/10.2166/wst.2020.443>.
82. McDonough LK, Santos IR, Andersen MS, O’Carroll DM, Rutledge H, Meredith K, et al. Changes in global groundwater organic carbon driven by climate change and urbanization. *Nat Commun.* 2020;11:1–10. <https://doi.org/10.1038/s41467-020-14946-1>.
83. Lawson CE, Nuijten GHL, de Graaf RM, Jacobson TB, Pabst M, Stevenson DM, et al. Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo ¹³C and ²H metabolic network mapping. *ISME J.* 2021;15:673–687. <https://doi.org/10.1038/s41396-020-00805-w>.

84. Kuenen JG. Anammox bacteria: From discovery to application. *Nat Rev Microbiol.* 2008;6:320–6.
85. Wang S, Radny D, Huang S, Zhuang L, Zhao S, Berg M, et al. Sci Rep. Nitrogen loss by anaerobic ammonium oxidation in unconfined aquifer soils. 2017;7:40173. <https://doi.org/10.1038/srep40173>.
86. Langone M, Yan J, Haaijer SCM, Op den Camp HJM, Jetten MSM, Andreottola G. Coexistence of nitrifying, anammox and denitrifying bacteria in a sequencing batch reactor. *Front Microbiol.* Frontiers Research Foundation; 2014;5:28. <https://doi.org/10.3389/fmicb.2014.00028>.
87. Yan J, Haaijer SCM, Op den Camp HJM, van Niftrik L, Stahl DA, Könneke M, et al. Mimicking the oxygen minimum zones: Stimulating interaction of aerobic archaeal and anaerobic bacterial ammonia oxidizers in a laboratory-scale model system. *Environ Microbiol.* 2012;14(12):3146–58. <https://doi.org/10.1111/j.1462-2920.2012.02894.x>.
88. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. *ISME J.* 2017;11:2399–406. <http://doi.org/10.1038/ismej.2017.113>.
89. Ali M, Oshiki M, Awata T, Isobe K, Kimura Z, Yoshikawa H, et al. Physiological characterization of anaerobic ammonium oxidizing bacterium “*Candidatus Jettenia caeni*.” *Environ Microbiol.* 2015;17(6):2172–89. <http://doi.org/10.1111/1462-2920.12674>.
90. Du MZ, Liu S, Zeng Z, Alemayehu LA, Wei W, Guo FB. Amino acid compositions contribute to the proteins’ evolution under the influence of their abundances and genomic GC content. *Sci Rep.* 2018;8:7382. <http://doi.org/10.1038/s41598-018-25364-1>.
91. Kartal B, Keltjens JT. Anammox Biochemistry: a Tale of Heme c Proteins. *Trends Biochem.* 2016;41(12):998–1011. <https://doi.org/10.1016/j.tibs.2016.08.015>.
92. Shimamura M, Nishiyama T, Shigetomo H, Toyomoto T, Kawahara Y, Furukawa K, et al. Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Appl Environ Microbiol.* 2007;73(4):1065–72. <https://doi.org/10.1128/AEM.01978-06>.

93. Maalcke WJ, Reimann J, De Vries S, Butt JN, Dietl A, Kip N, et al. Characterization of anammox hydrazine dehydrogenase, a Key -producing enzyme in the global nitrogen cycle. *J Biol Chem.* 2016;291(33):17077–92. <https://doi.org/10.1074/jbc.M116.735530>.
94. van de Vossenberg J, Woebken D, Maalcke WJ, Wessels HJCT, Dutilh BE, Kartal B, et al. The metagenome of the marine anammox bacterium “Candidatus” Scalindua profunda’ illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol.* 2013;15:1275–89. <http://doi.org/10.1111/j.1462-2920.2012.02774.x>.
95. Speth DR, Russ L, Kartal B, op den Camp HJM, Dutilh BE, Jetten MSM. Draft genome sequence of anammox bacterium “Candidatus Scalindua brodae,” obtained using differential coverage binning of sequencing data from two reactor enrichments. *Genome Announc.* 2015;3(1):e01415-14. <http://doi.org/10.1128/genomeA.01415-14>.
96. Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW, et al. Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature.* 2006;440(6):790–4. <http://doi.org/10.1038/nature04647>
97. Hira D, Toh H, Migita CT, Okubo H, Nishiyama T, Hattori M, et al. Anammox organism KSU-1 expresses a NirK-type copper-containing nitrite reductase instead of a NirS-type with cytochrome cd 1. *FEBS Lett.* 2012;586(11):1658–63. <http://doi.org/10.1016/j.febslet.2012.04.041>.
98. Narita Y, Zhang L, Kimura Z ichiro, Ali M, Fujii T, Okabe S. Enrichment and physiological characterization of an anaerobic ammonium-oxidizing bacterium “Candidatus Brocadia sapporoensis.” *Syst Appl Microbiol.* 2017;40(7):448–57. <http://doi.org/10.1016/j.syapm.2017.07.004>.
99. Speth DR, In’t Zandt MH, Guerrero-Cruz S, Dutilh BE, Jetten MSM. Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. *Nat Commun.* 2016;7:11172. <https://doi.org/10.1038/ncomms11172>.

100. Hu Z, Wessels HJCT, van Alen T, Jetten MSM, Kartal B. Nitric oxide-dependent anaerobic ammonium oxidation. *Nat Commun.* 2019;10:1–7.
<https://doi.org/10.1038/s41467-019-09268-w>.
101. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, et al. Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nat Microbiol.* 2021;6:354–365.
<https://doi.org/10.1038/s41564-020-00840-5>.
102. Ali M, Shaw DR, Albertsen M, Saikaly PE. Comparative genome-centric analysis of freshwater and marine anammox cultures suggests functional redundancy in nitrogen removal processes. *Front Microbiol.* 2020;11:1637.
<https://doi.org/10.3389/fmicb.2020.01637>.
103. Tang YP, Dallas MM, Malamy MH. Characterization of the BatI (Bacteroides aerotolerance) operon in Bacteroides fragilis: isolation of a B. fragilis mutant with reduced aerotolerance and impaired growth in in vivo model systems. *Mol Microbiol.* 1999;32(1):139–49. <http://doi.org/10.1046/j.1365-2958.1999.01337.x>.
104. Carmel-Harel O, Storz G. Roles of the glutathione- and thioredoxin-dependent reduction systems in the Escherichia coli and Saccharomyces cerevisiae responses to oxidative stress. *Annu Rev Microbiol.* 2000;54:439–61.
<http://doi.org/10.1146/annurev.micro.54.1.439>.
105. van Niftrik L, Geerts WJC, Van Donselaar EG, Humbel BM, Webb RI, Fuerst JA, et al. Linking ultrastructure and function in four genera of anaerobic ammonium-oxidizing bacteria: Cell plan, glycogen storage, and localization of cytochrome c proteins. *J Bacteriol.* 2008;190(2):708–17. <https://doi.org/10.1128/JB.01449-07>.
106. Castelle C, Guiral M, Malarte G, Ledgham F, Leroy G, Brugna M, et al. A new iron-oxidizing/O₂-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile Acidithiobacillus ferrooxidans. *J Biol Chem.* 2008;283(38):25803–11. <https://doi.org/10.1074/jbc.M802496200>.

107. Schouten S, Strous M, Kuypers MMM, Rijpstra WIC, Baas M, Schubert CJ, et al. Stable carbon isotopic fractionations associated with inorganic carbon fixation by anaerobic ammonium-oxidizing bacteria. *Appl Environ Microbiol.* 2004;70(6):3785–8. <https://doi.org/10.1128/AEM.70.6.3785-3788.2004>.
108. van de Vossenberg J, Woebken D, Maalcke WJ, Wessels HJCT, Dutilh BE, Kartal B, et al. The metagenome of the marine anammox bacterium “*Candidatus Scalindua profunda*” illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol.* 2013;15(5):1275–89. <https://doi.org/10.1111/j.1462-2920.2012.02774.x>.
109. van de Vossenberg J, Rattray JE, Geerts W, Kartal B, van Niftrik L, van Donselaar EG, et al. Enrichment and characterization of marine anammox bacteria associated with global nitrogen gas production. *Environ Microbiol.* 2008;10(11):3120–9. <http://doi.org/10.1111/j.1462-2920.2008.01643.x>.
110. Pedroni P, Volpe AD, Galli G, Mura GM, Pratesi C, Grandi G. Characterization of the locus encoding the [Ni-Fe] sulfhydrogenase from the archaeon *Pyrococcus furiosus*: evidence for a relationship to bacterial sulfite reductases. *Microbiology.* 1995;141:449–58. <https://doi.org/10.1099/13500872-141-2-449>.
111. Parey K, Demmer U, Warkentin E, Wynen A, Ermler U, Dahl C. Structural, biochemical and genetic characterization of dissimilatory ATP sulfurylase from *Allochromatium vinosum*. *PLoS One.* 2013;8(9):e74707. <https://doi.org/10.1371/journal.pone.0074707>.
112. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J.* 2016;10:761–77. <https://doi.org/10.1038/ismej.2015.153>.

Figure legends

Figure 1. Boxplot and stacked bar graph show the prevalence, relative abundance and diversity of anammox bacteria (*Brocadia*), based on 16S rRNA gene amplicon data. **a)** Number of OTUs identified as *Brocadia* relative to other bacteria (p:Other) and non-anammox *Planctomycetes* (p:Other *Planctomycetes*) across 80 samples. Horizontal line represents the median number of OTUs present. **b)** Structure of the anammox bacterial community observed in 60 samples where *Brocadia* were identified, ordered by DO concentration. Black points = DO content. Bars represent different genera within the class *Brocadia*. Acronyms: g = genus; f = family; o = order; c = class; p = phylum.

Figure 2. Diversity analysis and an ordination plot showing the environmental variables and factors influencing the structure of the anammox community. **a)** Boxplots represent observed OTUs (richness) and Inverse-Simpson diversity at anoxic, suboxic, dysoxic and oxic sites (Significance levels = $p < 0.01$; Wilcoxon Test). **b)** Distance-based Redundancy Analysis of the 16S amplicon data from the 60 groundwater samples using a Bray-Curtis dissimilarity matrix between samples based on OTU abundance. Vectors show significant environmental variables (excluding ORP due to missing values), constraining the variability in community composition. *Brocadia* OTUs were added to the ordination using species scores (shown as large coloured squares) and the 5 OTUs shown are identified to the lowest level of taxonomy (Acronyms: g = genus; f = family; o = order; c = class; p = phylum).

Figure 3. Correlations between 16S rRNA gene amplicon data, aerobic ammonia oxidisers (AOA and AOB) and abundance of hydrazine synthase genes and transcripts. **a)** Correlation of *hzsB* gene per L of groundwater (log10) and number of *Brocadia* (log10) sequences within the rarefied OTU table with samples containing class *Brocadia*. **b)** Boxplot representing log10 abundance of *hzsB* genes for genes and transcripts. Black open circle represents the mean and horizontal line represents the median of gene copies per L. Purple gradient for **a)** and **b)** correspond with DO concentration (mg/L). **c)** Correlation of *amoA* genes (log10) and *hzsB* genes (log10) copies per L of groundwater using Spearman correlation for genetic potential

(DNA). **d)** Correlation of *amoA* genes (log10) and *hzsB* genes (log10) copies per L of groundwater using Spearman correlation in transcripts (RNA).

Figure 4. Phylogenetic distribution and relative abundance of *Brocadia* MAGs. **a)** Maximum-likelihood phylogenomic tree of 28 *Planctomycetes* genomes based on 120 concatenated bacterial marker genes (GTDB-Tk) with 5,040 amino-acid sites using LG+F+R5 model of substitution and 1,000 bootstraps. Purple gradient represents the DO concentration (mg/L) at the site with highest relative genome abundance for that genome. Coloured tiles represent the environment of recovery for the reference genomes. Bootstrap values shown as black circles equal 100%. Scale bar indicates the number of substitutions per site. Sequences from this study are shown in bold font, with both the study identifier and GTDB classification given. Orange colour strip represents clades of *Brocadia* genomes recently recovered from aquifers. See Table S5 for reference genome details. **b)** *Brocadia* genome relative abundance across DO concentration (mg/L) at each site (relative to other *Brocadia*). Bubble size corresponds to nitrate concentration (g/m³) at each site.

Figure 5. Plot of dissolved nitrogen versus dissolved argon concentrations at sites sampled for metagenomics analysis. Dissolved Ar, N₂ are expressed in mL of the respective gas at standard temperature and pressure (STP=273.15 K, 101.325 kPa) per kg of water. Bold line represents gas concentrations in water which are in equilibrium with the atmosphere at the given temperature. Arrows indicate competing processes that can alter gas concentrations, where grey dashed lines indicate excess air in groundwater relative to atmosphere (with upper and lower lines representing addition of unfractionated excess air relative to equilibrium concentrations at 10 and 15°C). Black horizontal arrow depicts additional excess N₂ inferred to be from biological processes (denitrification or anammox). Reconstructed N₂ data (in equilibrium with inert atmospheric gases), based on groundwater recharge temperatures and excess air concentrations derived from dissolved Ne and Ar data (shown as red hollow symbols). Recharge temperature = temperature of recharging water at the time it enters the groundwater system. Excess air = dissolved air in excess of the equilibrium soluble amount at

the given recharge temperature (thought to originate from processes such as bubble entrapment occurring during recharge and subsequent dissolution under increased hydrostatic pressure). The difference between these and the measured N₂ data (blue solid symbols, and shift their along x-axis relative to red hollow symbols) indicates the amount of N₂ in excess, formed via denitrification and/or anammox at each site. Error bars show the combined statistical standard uncertainty from all processes and calculations contributing to the measurement uncertainty, expressed as one standard deviation. Groundwater from wells SR1-2, BW8, BW19, RF2-3 is characterised as oxic, while groundwater from E1 and N3 is dysoxic-suboxic (Table S1).

Figure 6. An overview of the predicted metabolic pathways of characterised anammox bacteria built from the summarised annotated data from the reconstructed genomes and mapped transcripts. **a)** Metabolic pathways present in MAGs. Nxr, nitrite:nitrate oxidoreductase; Nir, nitrite reductase; Nrf, nitrite reductase forming ammonium; HZS, hydrazine synthase; HDH, hydrazine dehydrogenase; AmtB, ammonium transporters; FocA, nitrite transporters; NarK, nitrite/nitrate transporter; ETM, electron transfer module from quinone pool to HZS (composed of kuste2856 and kuste2855); R/b, Rieske/cytochrome b (bc₁) complexes, R/b-2 (kustd1480-85), and R/b-3 (kuste4569-74); F-type ATPase, F-type ATP synthase (MAGs containing ≥50% of subunits); GAP, Glyceraldehyde 3-phosphate; EMP, Embden-Meyerhof-Parnas pathway; FDH, Formate dehydrogenase; FHS, Formate--tetrahydrofolate ligase; FOLD, methylenetetrahydrofolate dehydrogenase; MetF, Methylenetetrahydrofolate reductase; AcsE, 5-methyltetrahydrofolate:corrinoid; AcsB, acetyl-CoA synthase; AcsA acetyl-CoA synthetase; PPP, Pentose phosphate pathway; TCA, tricarboxylic acid cycle. Colours represent genomes from this study, number equals copies present.* = Partial HzsA subunit. **b)** Logged (log 10) TPM values for active hydrazine synthase, hydrazine dehydrogenase and nitrite reductase forming ammonium genes from groundwater characterised as oxic (samples gwj09 from well SR1 and gwj11 from well SR2) and dysoxic

696 (gwj13-14 from well E1 and gwj15-16 from well N3). Samples gwj14, gwj16 are of sonicated
697 groundwater.

698 **Figure 7.** Phylogenetic tree of the recovered hydrazine synthase protein subunits from
699 *Brocadia*. **a)** Phylogenetic tree with 65 protein sequences of hydrazine synthase subunits
700 with 964 amino-acid sites, and built using WAG+G4 model using 1,000 bootstraps. The
701 hydrazine synthase (HZS) alpha, beta and gamma proteins were predicted from protein coding
702 sequences recovered from genomes in the study (red) and other HZS protein sequences
703 available from Uniprot database. HDH = Hydrazine dehydrogenase. Bold black sequences
704 are from representative model organism *Ca. Kuenenia stuttgartiensis*. Circles represent > 50%
705 bootstrap value. Scale bar represents number of substitutions per site. **b)** Phylogenetic tree of
706 fused Hzs beta-gamma proteins (from predicted protein coding sequences) made with 10
707 sequences (6 HzsA subunits) and 903 amino-acid sites, and built using WAG+G4 model
708 (1,000 bootstraps). Circles represent > 75% bootstrap value, red sequence was recovered in
709 this study.

710

a)

b)

a)

b)

a)

b)

c)

d)

a)

Tree scale: 0.1

b)

b)

a)

b)