Genomic profiling of *Nitrospira* species reveals ecological success of comammox *Nitrospira*

Alejandro Palomo1,2*, Arnaud Dechesne1, Anders G. Pedersen3 and Barth F. Smets1

**Abstract**

**Background:** The discovery of microorganisms capable of complete ammonia oxidation to nitrate (comammox) has prompted a paradigm shift in our understanding of nitrification, an essential process in N cycling, hitherto considered to require both ammonia oxidizing and nitrite oxidizing microorganisms. This intriguing metabolism is unique to the genus *Nitrospira*, a diverse taxon previously known to only contain canonical nitrite oxidizers. Comammox *Nitrospira* have been detected in diverse environments; however, a global view of the distribution, abundance, and diversity of *Nitrospira* species is still incomplete.

**Results:** In this study, we retrieved 55 metagenome-assembled *Nitrospira* genomes (MAGs) from newly obtained and publicly available metagenomes. Combined with publicly available MAGs, this constitutes the largest *Nitrospira* genome database to date with 205 MAGs, representing 132 putative species, most without cultivated representatives. Mapping of metagenomic sequencing reads from various environments against this database enabled an analysis of the distribution and habitat preferences of *Nitrospira* species. Comammox *Nitrospira*'s ecological success is evident as they outnumber and present higher species-level richness than canonical *Nitrospira* in all environments examined, except for marine and wastewaters samples. The type of environment governs *Nitrospira* species distribution, without large-scale biogeographical signal. We found that closely related *Nitrospira* species tend to occupy the same habitats, and that this phylogenetic signal in habitat preference is stronger for canonical *Nitrospira* species. Comammox *Nitrospira* eco-evolutionary history is more complex, with subclades achieving rapid niche divergence via horizontal transfer of genes, including the gene encoding hydroxylamine oxidoreductase, a key enzyme in nitrification.

**Conclusions:** Our study expands the genomic inventory of the *Nitrospira* genus, exposes the ecological success of complete ammonia oxidizers within a wide range of habitats, identifies the habitat preferences of (sub)lineages of canonical and comammox *Nitrospira* species, and proposes that horizontal transfer of genes involved in nitrification is linked to niche separation within a sublineage of comammox *Nitrospira*.

**Background**

The biological oxidation of ammonia to nitrate via nitrite, termed nitrification, is an essential process in terrestrial and aquatic environments. Nitrification links oxidized and reduced pools of inorganic nitrogen, contributes to nitrogen loss from agricultural soils reducing fertilization efficiency [1], is a crucial process in water quality engineering [2], and can lead to the production of nitrous oxide, a strong greenhouse gas [3]. For more than one century, this central process in the nitrogen cycle was assumed to be a two-step process catalyzed by two distinct functional groups: ammonia-oxidizing microorganisms (AOM—consisting of bacteria and archaea) and
nitrite-oxidizing bacteria (NOB). In 2015, microorganisms capable of the complete oxidation of ammonia to nitrate (complete ammonia oxidation aka comammox) were discovered [4, 5] within the *Nitrospira* genus. This genus was previously considered to only harbor nitrite oxidizers (here also referred to as canonical *Nitrospira*) capable of using alternative metabolisms such as hydrogen or formate oxidation, but not ammonia oxidation [6].

Phylogenetic evidence indicates a complex evolutionary history of comammox capability and it is currently unclear whether or not it was present in ancestral *Nitrospira* [7]. Clearly, the gain (or the loss) of such a trait must have had strong ecological consequences. Indeed, comammox *Nitrospira* can capture a much greater amount of energy from ammonia oxidation than canonical *Nitrospira* from nitrite oxidation, but most likely at the cost of a lower maximum growth rate [8, 9]. In addition, comammox and canonical *Nitrospira* directly compete with different guilds of nitrifiers (AOM and NOB, respectively), resulting in a very different selective landscape. Yet, we have little understanding of how these markedly different metabolic strategies affect *Nitrospira*'s current ecological distribution and how this distribution relates to its evolutionary history.

Phylogenetic diversity within *Nitrospira* is high [6]: the genus consists of at least six lineages with pronounced divergence at the 16S rRNA level (sequence similarities < 90%). Its members are environmentally widespread [6], and are found in soils [10], freshwaters [11], oceans [12], groundwaters [13], and technical systems [14, 15]. Comammox *Nitrospira* species have been detected in different natural environments such as soils, rivers, lakes or coastal wetlands [16–19]; as well as in engineered systems such as waste and drinking water treatment plants [4, 20]. Estimates from a limited number of sites indicate that comammox *Nitrospira* is abundant in drinking water treatment plants (DWTP) [21], and to a lesser extent, in wastewater treatment plants (WWTP) [22] and soils [23, 24]. So far, the distribution or abundance estimates of comammox *Nitrospira* have been based on either functional gene (subunit A of the ammonia monooxygenase (amoA)) amplicon-based profiling [16, 25, 26] or on limited genome-based profiles in specific environments [23, 27–29]. As a consequence, our view of the distribution, abundance, and diversity of *Nitrospira* is fragmented and likely incomplete. shotgun metagenomic sequencing profiling has been successfully used to disclose the ecological patterns of various microbial populations at the large scale [30–32]. In this study, we used metagenome assembled genomes (MAGs) to create the largest *Nitrospira* genomes database to date. By comparing the raw sequencing reads in metagenomic data to these genomes (“read recruitment”), we determined the presence and abundance of *Nitrospira* species in various environments across the globe and elucidated how environment and spatial distance affect *Nitrospira* distribution, and to what degree *Nitrospira* phylogeny associates with its ecology. Our analysis provides a global survey of *Nitrospira* distribution and abundance with species-level resolution and unravels the niche preferences of the different comammox *Nitrospira* types.

**Results**

**Genomic characteristics of *Nitrospira* genomes**

Using a combination of multiple automatic binning tools followed by several refinement steps to improve the bin quality, we retrieved 55 metagenome-assembled *Nitrospira* genomes (MAGs) from newly obtained and publicly available metagenomes (Table S1). In addition, we downloaded *Nitrospira* MAGs from public databases (up to January 2022), resulting in a database of 205 MAGs from drinking water treatment plants (DWTP) (*n* = 29), freshwater (*n* = 45), groundwater (*n* = 10), marine (*n* = 14), soil (*n* = 11), wastewater treatment plants (WWTP) (*n* = 92) and other environments (*n* = 4); from across the globe (Canada, *n* = 12; China, *n* = 58; Europe, *n* = 61; Oceania, *n* = 12; USA, *n* = 37; others, *n* = 25) (Table S2). The phylogenomic analysis of the 205 MAGs placed 48 MAGs into *Nitrospira* lineage I, 121 into lineage II, 27 into lineage IV, and 9 into other lineages (Fig. 1). Two genomes recovered from groundwater samples (*Nitrospira* sp. H14_bin041 and *Nitrospira* sp. H32_bin031) not associated to known lineages and phylogenetically distant to other nitrifying-*Nitrospira* spp. were included in the catalog as they both contain genes involved in nitrite oxidation (*nxrA* and *nxrB*). The 205 MAGs span 132 putative species (further on simply referred to as ‘species’) using a threshold average nucleotide identity (ANI) of ≥ 96% (Figure S1 and Table S2). Of the 205 MAGs, 93 are comammox *Nitrospira* (69 clade A and 24 clade B) (Fig. 1 and Table S1). Similarly, out of the 132 *Nitrospira* species, 66 are comammox *Nitrospira* (49 clade A and 17 clade B) (Fig. 1). Average genome completeness and contamination were estimated at 89% (70 to 98%) and 1% (0 to 5%), respectively (Table S2). *Nitrospira* MAGs have an average genome size of 3,435,709 ± 860,289 bp, and contain 3541 ± 844 genes. Although there was no significant difference in the genome completeness across different *Nitrospira* lineages, there were obvious dissimilarities in the genome size and gene number (Table 1). The variability in genome size is marked within Lineage IV, with one branch (IV type A), associated with MAGs from wastewater treatment plants, characterized by genomes much larger than those ones in the second branch, dominated by MAGs from marine/sponges (IV type B). Thus, genomes in the
lineage IV type B and genomes belonging to two lineages distinct from lineages I, II, and IV (referred to as 'other lineages') are significantly smaller and harbor less genes than other Nitrospira genomes ($p < 0.05$). The GC content of the genomes is 56.2% ± 2.8. The lowest GC content was detected in the lineage IV type A genomes (49.8% ± 0.9), while lineage I (58.8% ± 0.6) and genomes assigned to other lineages have highest GC (62.0% ± 1.3) (Table 1). Interestingly, the two clearly separated groups within Lineage IV (A and B) are significantly different in genome size, number of genes and GC content (Table 1 and Fig. 1).
Comparison of Nitrospira genomes

A pangenomic analysis of 205 Nitrospira MAGs grouped a total of 723,029 coding sequences (CDS) into 21,450 gene clusters (GCs), with a Nitrospira core genome containing 1190 GCs (Table S3), a similar number and metabolic content to our previous study of 16 Nitrospira genomes [7] (59,744 CDS grouped into 12,337 GCs, with a core genome consisting of 1382 GCs). The core genome includes genes for the nitrite oxidation pathway, the reductive tricarboxylic acid cycle for CO₂ fixation (rTCA), gluconeogenesis, nitrite oxidation pathway, the reductive tricarboxylic cycle. Chlorite dismutase and copper-containing nitrite reductase (nirK) are also present in the core genome. We identified a total of 49 comammox-specific GCs; 18 and 3 of these GCs have highest sequence similarity to homologs in betaproteobacterial ammonia-oxidizing bacteria (AOB) and gammaproteobacterial methane oxidizers, respectively (Table S3). These genes encode the ammonia oxidation pathway, as well as specific urea transporters and copper homeostasis proteins (Table S3). We identified 47 and 25 comammox clade A and clade B-specific GCs, respectively, mostly with unknown function (Table S3). In addition, each Nitrospira genome harbors an average of 161 unique gene clusters, most of unknown function.

Comammox Nitrospira spp. are widely distributed and often dominant over canonical Nitrospira

We characterized the distribution of the 132 Nitrospira species in 788 metagenomes across eight broadly defined environments (drinking water treatment plant (DWTP), n = 40; freshwater, n = 167; groundwater, n = 86; hot spring, n = 36; marine, n = 75; soil, n = 206; wastewater treatment plant (WWTP), n = 177; others, n = 1) across the globe (China, n = 139; Europe, n = 128; USA, n = 309; others, n = 212) (Table S4). Nitrospira species were detected in 598 metagenomes (Fig. 2). The most widely distributed Nitrospira species in the investigated metagenomes were the canonical lineage I Nitrospira sp. ND1 (frequency of occurrence of 18%) and Nitrospira sp. UBA10644 (16%), and the clade B comammox Nitrospira sp. LM_bin98 (16%) (Figure S2). Forty-one species were found in less than 1% of the metagenomes, including several of the Nitrospira species with a representative isolate or enrichment such as Nitrospira sp. KM1, N. japonica, N. marina (Figure S2).

All the DWTP (40/40), most of the WWTP (147/177), groundwater (63/86), freshwater (127/167), and soil metagenomes (180/206); and about one third of the hot spring (13/36) and marine metagenomes (27/75) have at least one Nitrospira species. The Nitrospira species richness per metagenome is highest in the DWTP metagenomes (10.0 ± 6.7), followed by groundwater (5.3 ± 5.4), WWTP (4.1 ± 3.5) and freshwater (2.9 ± 2.9), while it is lowest in the hot spring (1.4 ± 0.7), marine (1.4 ± 0.7) and soil metagenomes (1.1 ± 0.4). The average number of detected comammox Nitrospira species exceeds that of non-comammox species in DWTP, freshwater, and hot springs, while the opposite was observed for WWTP and marine samples (P < 0.05), where comammox was not detected (Fig. 3A). No significant difference was found in groundwater metagenomes (Fig. 3A). The DWTP metagenomes have the highest abundance of Nitrospira spp., while hot spring and soil metagenomes have the lowest abundance of Nitrospira MAGs (Fig. 3B). Comammox Nitrospira constitute the majority of Nitrospira species in DWTP, hot spring, and soil metagenomes, while canonical species are significantly more abundant than comammox Nitrospira in marine and WWTP samples (P < 0.05) (Fig. 3B).

Environment determines Nitrospira spp. distribution, without large-scale biogeographical signal

To examine the distribution pattern of Nitrospira species, we conducted a principal component analysis...
(PCA) of the relative abundance of *Nitrospira* species in the metagenomes. WWTP metagenomes clearly cluster and separate from DWTP metagenomes (Fig. 4A); soil metagenomes also cluster together, while groundwater and freshwater metagenomes have more variable *Nitrospira* compositions (Fig. 4A).
Within each habitat, we found, in most cases, a weak, but significant correlation between geographical distance and *Nitrospira* community dissimilarity (Figure S3). However, these correlations nearly disappear when samples within small distances were excluded from the analysis, with the exception of DWTP (Figure S3).

Analysis of the *Nitrospira* species abundances across the metagenomes showed that *Nitrospira* spp. belonging to the same (sub)lineages tend to co-occur, and also, revealed the existence of niche separation between *Nitrospira* (sub)lineages (Fig. 4B, Figure S4 and Figure S5). Lineage I species are primarily found in WWTP metagenomes while the distribution of lineage II species is more varied. Clade A comammox *Nitrospira* species are distributed in two clusters (Fig. 4B, Figure S4). Some of the clade A species co-occur with the *Nitrospira* lineage I species in WWTP samples, while the other clade A species are more typical of DWTP along with some of the groundwater and freshwater metagenomes, and co-occur with the lineage I *Nitrospira lenta*-type species. The clade B comammox *Nitrospira* species are frequently present in DWTP, groundwater, freshwater, and soil metagenomes (Fig. 4B and Figure S4). Lineage IV species are characteristic of marine samples, while *Nitrospira inopinata* is the dominant species present in hot springs (Fig. 4B and Figure S4).

**Relation between ecological and phylogenetic similarities**

We assessed which feature better explained habitat preference by comparing, for pairs of species, similarity in their environmental distribution (as measured by the correlation of their abundance across samples) with overall genomic similarity (as measured by the average amino acid identity between their genomes). Similarity in environmental distribution was also compared to sequence similarity of the key nitrification enzymes ammonia monooxygenase (AMO), hydroxylamine dehydrogenase (HAO) and nitrite oxidoreductase (NXR). We found that genomic similarity between pairs of species is positively correlated to similarity in habitat preference (Mantel statistic \( r = 0.47, P < 0.001 \)) (Table 2). A similar correlation was obtained when only comammox species or canonical species were evaluated (\( r = 0.63, P < 0.001; r = 0.69, P < 0.001 \)). This correlation was especially strong when only the canonical *Nitrospira* species from lineage I and II were considered (\( r = 0.83, P < 0.001 \)). Similar trends as for AAI were observed when using a set of 120 translated single-copy genes, although with slightly weaker correlation values (Table 2). Regarding the enzymes involved in nitrification, we observed very weak correlations between NxrB similarity, considered a powerful functional and phylogenetic marker [6], and habitat preference with the exception of canonical *Nitrospira* species affiliated to lineage I and II (\( r = 0.70, P < 0.001 \)). The correlation between AmoA sequence similarity and habitat...
similarity was weaker than for whole genome similarity for comammox *Nitrospira* species (*r* = 0.46, *P* < 0.001). HaoA similarity, however, displayed a stronger correlation with habitat similarity (*r* = 0.60, *P* < 0.001).

**Hydroxylamine dehydrogenase is potentially involved in recent niche separation within clade A**

These results indicate that the hydroxylamine dehydrogenase might be a better indicator than other nitrification enzymes of niche preference for comammox *Nitrospira*. Consistent with this, phylogenetic analysis of HaoA showed a division of clade A into two subclusters (referred to as clade A1 and clade A2, respectively) similar to the one identified from habitat preference patterns (Fig. 4 and Figure S4). Indeed, clade A2 sequences are clearly separated from those of clade A1, and more closely related to some of the clade B species (Fig. 5). This contrasts with the clear monophyletic separation of the two comammox clades (A and B) based on phylogeny of AmoA (Fig. 5), which is also supported by the phylogenetic analyses of the concatenation of 120 translated single-copy genes (Fig. 1). Further, we detected two *nirK* genes (encoding a copper-containing nitrite reductase) next to the HAO cluster in the clade A2 genomes. This gene synteny was also found in several of the clade B genomes, but never in the clade A1 genomes (Figure S6). This shared synteny of the HAO genetic region suggests a horizontal gene transfer event between clade A2 and clade B (as also proposed in our earlier evolutionary analysis [7]) instead of convergent evolution of the hydroxylamine dehydrogenase.

**Discussion**

This study represents the first effort to analyze the global distribution and ecological niches of *Nitrospira* species, including comammox *Nitrospira*. We combined a large number of *Nitrospira* MAGs retrieved in this study with publicly available genomes to build a dataset consisting of 205 medium to high quality *Nitrospira* genomes (> 70% completeness and < 5% contamination). These genomes were recovered from very different environments and geographical areas, and spanned five different lineages. This dataset properly covers the most ubiquitous and species-rich lineages (I and II) as well as both comammox clades. It also provides a unique insight into the phylogenomics of lineage IV which demonstrates that this deep-branching lineage experienced stronger genomic changes than the rest of the lineages and should probably be considered as comprising of two separate lineages, as previously proposed from 16S rRNA gene sequence analysis [33]), with starkly contrasting genomic features (genome size, gene number, and GC content). In addition, genomes from other lineages lacking cultivated representatives were present in the database. On the other hand, no MAGs from the previously established lineages V and VI were represented in the database. Nevertheless, only one enrichment (*Candidatus Nitrospira bockiana*; lineage V), one isolate (*Nitrospira calida*; lineage VI) and few 16S rRNA sequences have been recovered from these lineages that are believed to occupy very specific habitats [6, 34, 35]. Overall, the database built in this study represents the larger genomic dataset used to study the genomic characteristics and ecology of the *Nitrospira* genus.

We quantified the relative abundance and distribution of *Nitrospira* species in different habitats by exploiting large amounts of shotgun metagenomic sequencing data. Thus, we mapped metagenomic sequencing reads against a *Nitrospira* database of 15 universal single-copy genes from the *Nitrospira* species retrieved in our study [31]. The uncovered diversity of *Nitrospira* species, and specifically the vast diversity within lineage II *Nitrospira*, reflects the species-level diversity previously estimated from 16S rRNA sequences [6]. Our analysis confirms the previously described ubiquity of *Nitrospira* [6] and reveals the extent of the ecological success of comammox *Nitrospira*. We detected comammox species within a wide range of climatic zones, from polar (soils in Svalbard and Antarctica) to tropical (soils in Vietnam, and fresh and wastewater in Singapore) and temperate zones. We found that comammox *Nitrospira* coexist with and are more abundant than canonical *Nitrospira* in all studied habitats except in WWTP and marine samples. This hitherto unrecognized dominance of comammox

| Marker      | Whole genome | Single-copy genes | AmoA | HaoA | NxrB |
|-------------|--------------|-------------------|------|------|------|
| Nitrospira  | 0.47***      | 0.34***           | –    | –    | 0.08 |
| Comammox    | 0.69***      | 0.57***           | 0.46*** | 0.60*** | 0.23** |
| Canonical   | 0.63***      | 0.52***           | –    | –    | 0.10 |
| Canonical (Lineage I and II) | 0.83*** | 0.71***         | –    | –    | 0.70*** |

*p < 0.05; **p < 0.01; ***p < 0.001*
Nitrospira implies that, until recently, by equating detection of Nitrospira 16S rRNA or nxrB with strict NOB metabolism, the number of ammonia oxidizers in the environment has been systematically and significantly underestimated. Among the studied habitats, comammox Nitrospira is especially diverse and abundant in drinking water treatment systems, which are characterized by surface-attached microbial communities and low ammonium fluxes [15, 36], a suitable environment for high growth yield microorganisms such as comammox Nitrospira [8, 37]. On the other hand, canonical Nitrospira clearly outnumber comammox Nitrospira in WWTP. These results are line with a study conducted on samples collected from hundreds of WWTP across the world, which shows canonical Nitrospira lineage I species as most abundant Nitrospira [38]. This predominance of canonical nitrite oxidizers in WWTP is consistent with the proposition that nitrogen-rich environments select for division of labour because it maximizes growth rates [8, 39]. However, in some full-scale WWTP [22, 40] and lab-scale reactors [41], comammox Nitrospira have been found more abundant than canonical ammonia oxidizers, suggesting that factors beyond nitrogen fluxes influence comammox Nitrospira fitness in these systems. Based on a survey of 14 WWTP, Cotto et al. (2020) concluded that long solids retention time favor the prevalence of comammox bacteria [42]. Nitrospira MAGs were detected in marine metagenomes, although in most cases, they were samples collected from coral reef or/and sponges. This is consistent with Nitrospira being a minor nitrite oxidizing bacterium in oceans [6]. Comammox were not detected in the marine metagenomes, as all Nitrospira species identified in this habitat were restricted to lineage IV. Previously

---

**Fig. 5** Maximum likelihood phylogenetic trees of Nitrospira spp. based on (left) AmoA and (right) HaoA amino acid sequences. Only complete or nearly complete sequences are included. The strength of support for internal nodes was assessed by performing bootstrap replicates, with the obtained values shown as colored circles (top left legend). The comammox Nitrospira clades and subclades are indicated by colored boxes, and the color font of clade A sequences in the AmoA tree refers to their membership on the HaoA tree. Gamma- and betaproteobacterial AOB sequences were used to root the trees. Shapes next to the sequence names denote the environment from where the MAGs were recovered (color code as in Fig. 1)
studies have detected comammox in estuaries [43, 44] and coastal sediments [45], but consistently with our findings, no complete ammonia oxidizers have been found in open ocean water. Multiple studies have highlighted the relevance of comammox Nitrospira spp. in different soil systems [19, 46–49], however, our analysis showed relatively low presence, abundance and richness of Nitrospira spp. in soil metagenomes. This habitat generally contains a large diversity of microorganisms, and recovery of MAGs from this kind of complex metagenomes can be challenging. Therefore, it might be possible that Nitrospira species from soils are underrepresented in our dataset. Based on amoA amplicon sequencing a variety of comammox Nitrospira, especially assigned to clade A, have been detected in soils [19, 47, 50]. This contrasts with the limited number of comammox Nitrospira clade A MAGs recovered from soils metagenomes in our catalog. To circumvent this potential underrepresentation, we extended our analysis beyond the species level (ANI ≥ 95%) by mapping metagenomics reads to our database using an ANI cutoff of 75% (genus level). The presence and number of Nitrospira spp. detected in the soil metagenomes increased, however, richness and abundance were still lower than in most other environments (Table S5). Nevertheless, efforts to recover more genomes from soil environments or conducting global-scale studies using others approaches (such as qPCR) would help to confirm our observations.

In contrast to what has been observed for other bacterial taxa such as Streptomycetes [51], SAR11 [52] or Acidobacteria [53], we did not detect a strong biogeographical signal for Nitrospira spp., with the exception of drinking water treatment plants and, to a lesser extent, wastewater treatment plants, where dissimilarity of the Nitrospira communities correlates with geographical distance. However, due to the limited number of DWTP samples, a larger dataset is needed to confirm this observation. For all other habitats, the correlation nearly disappears after few tens of kilometers, which might correspond to the limit of effective dispersal of Nitrospira species or might reflect the scale of variability of environmental parameters that affect Nitrospira spp. Environmental filtering controls Nitrospira species distribution and, since closely related species tend to share habitats preference, it results in a clear phylogenetic clustering of Nitrospira types in environmental communities. As observed for other microbial taxa [54, 55], dissimilarity in habitat preference generally increases with genomic dissimilarity, with the strongest correlation for canonical Nitrospira species. This pattern is consistent with progressive phylogenetic and ecological divergences, as expected from genetic drift or from random fluctuation in natural selection [56]. Yet, this pattern of congruency between genetic similarity and habitat similarity is not always true when considering the genes essential to nitrification. Indeed, the similarity in NxrB sequences, used as functional and phylogenetic markers for nitrite oxidation [10], does not predict habitat similarity for comammox Nitrospira. In contrast, HaoA sequence similarity strongly correlates with comammox Nitrospira habitat preference. This is because HaoA phylogeny matches the striking niche separation we observed within clade A comammox Nitrospira species. While clade A2 species were mostly found in DWTP, groundwater and freshwater sharing habitats with clade B comammox species, comammox clade A1 species primarily occupy the habitat typical of lineage I Nitrospira species (WWTP). This niche separation is not captured by amoA phylogeny and, as result, the distinction of two subclades of clade A comammox amoA (also named A1 and A2 [25]), provides less accurate information on habitat preference. This can be illustrated by the fact that amoA-based clade A1 has both representatives that frequently occur in WWTP (N. nitrosa) and others (Nitrospira sp. RF51 cluster) that are typical of DWTP, groundwater and freshwater samples (Fig. 4 and Figure S4).

Ecological divergence of closely related types can emerge from gene acquisition through horizontal gene transfer [57–59] or by changes in existing genes whose divergence can be accelerated by natural selection [60–62]. The high sequence divergence for HAO subunits between clade A1 and clade A2 genomes, and the similarity between clade A2 HAO and clade B HAO is unlikely to result from convergent evolution, and horizontal gene transfer is thus the likely cause of the ecological divergence between clade A1 and clade A2 comammox Nitrospira. Indeed, clade A2 and some of the clade B genomes share a unique hao syntenic that includes a duplicate nirK, which is absent in clade A1 genomes. Such transfer event between clade A2 and B had been proposed in a previous, more limited evolutionary analysis [7]. It is plausible that this transfer would provoke a niche modification in the recipient as HAO is one of the key enzymes for ammonia oxidation [3]. The apparent co-transfer of adjacent nirK genes is intriguing, as NirK has recently been posited as essential in ammonia oxidation [63]. Nevertheless, although our analysis shows that the genomes involved in this proposed horizontal gene transfer tend to share similar habitats, further ecophysiological analyses are needed to confirm the basis of the ecological divergence within clade A. This will require continuous efforts at isolating representatives of all comammox Nitrospira (sub)clades, beyond the few currently isolate and enrichments, which belong to comammox clade A1 [5, 37, 64].
Conclusions
This study represents the most extensive survey of the Nitrospira genus at genome level performed to date. Complete ammonia oxidizers are present in a large variety of habitats, biomes and global regions, and often dominate over canonical Nitrospira. Our analysis suggests that the type of environment governs Nitrospira species distribution, without large-scale biogeographical signal. We identified the habitat preferences of (sub) lineages of canonical and comammox Nitrospira species, and proposed that horizontal transfer of genes involved in nitrification could be linked to niche separation within a sublineage of comammox Nitrospira. Together, these findings provide deeper insights into the ecology of an important player of the biogeochemical nitrogen cycle. Future studies involving canonical ammonia oxidizers are needed to determining the functional redundancy and niche differentiation of comammox, ammonia oxidizing bacteria and ammonia oxidizing archaea at global scale.

Methods
Sample collection and DNA extraction
Filter material (15 ml) was collected from 2 locations or from 1 location at two different times at the top of the filters of 12 Danish waterworks using a 1% hypochlorite-wiped stainless-steel grab sampler. Filter material was immediately placed into cryotubes, immersed in liquid nitrogen and stored at –80°C for further analysis. DNA was extracted from 0.5 g of sand material using the MP FastDNA Spin Kit (MP Biomedicals LLC, Solon, USA) as described elsewhere [15].

Library preparation and sequencing
DNA libraries were generated from the 24 extracted DNA samples with the Nextera XT DNA library preparation kit (Illumina Inc.) according to the manufacturer’s instructions. The samples were sequenced in one lane, on a HiSeq 4000, 150 bp pair-end with dual indexing at BGI, Copenhagen.

Recovery and assessment of metagenomic assembled genomes
Trimmomatic v0.22 [65] was used to remove adapters and trim the reads (threshold quality = 15; minimum length = 40). Quality control was carried out using FastQC (Babraham Bioinformatics (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). High-quality reads from each sample were co-assembled into scaffolds using IDBA-UD [66] with the options --pre_correction--min_contig 1000. In addition, 24 single-sample assemblies were performed following the same procedure. Metagenomic binning and refinement approach was conducted as previously described [67], and the overall workflow of the recovery of metagenomic assembled genomes is visualized in Figure S7.

In addition, metagenomic binning from several metagenomes downloaded from public databases (Table S1) was carried out using the binning, refinement and reassembly modules of metaWRAP [68]. Trimming and quality control of the raw reads as well as de novo assemblies of these metagenomes were carried out as described above.

Database construction
Fifty-five Nitrospira genomes recovered as part of this study and publicly available Nitrospira genomes were included in the genome dataset. Completeness and contamination rates of the all final population bins were assessed using CheckM [69]. A final set of 205 MAGs with completeness ≥70% and contamination ≤5% were included in the database.

For the metagenomic datasets, raw reads from 788 metagenomes downloaded (on January 2022) from NCBI [70] and MG-RAST [71] were included in the analysis. These metagenomes were selected as follows: The NCBI Sequence Read Archives (SRA) (sequencing type “whole genome sequencing”, “HiSeq”, and environmental package: “air”, “aquifer”, “biofilm”, “biofilter”, “coral reef”, “estuary”, “freshwater”, “groundwater”, “hot spring”, “marine”, “metagenome”, “permafrost”, “rhizosphere”, “rice paddy”, “sand”, “sediment”, “soil”, “sponge” “urban”, “wastewater”, “wetland”) were queried with amoA and nrxB sequences from several Nitrospira spp. to identify datasets most likely to contain sequences associated with Nitrospira spp. Metagenomic datasets identified as possessing Nitrospira spp.-associated sequences were included in the metagenomic dataset. In addition, MG-RAST metagenomes not present in NCBI SRA, and with more than 2000 reads taxonomically annotated as Nitrospira were included in the metagenomic dataset. No more than six samples from the same study were included in the final metagenomic dataset used in this study to avoid their overrepresentation. Trimming and quality control of the raw reads as well as de novo assemblies were carried out as described above.

Species abundance estimation
A 96% average nucleotide identity (ANI) cutoff was used to cluster genomes into groups of species as we identified a discontinuity in pairwise ANI for the 205 Nitrospira MAGs (Figure S8). Therefore, the Nitrospira genomes were dereplicated using dRep with the secondary clustering threshold set at 96% gANI. Among the genomes classified as belonging to the same species, the one with highest quality was chosen as representative for that species. The species abundance of
each representative genome across the studied metagenomes was assessed using MIDAS [31]. Briefly, MIDAS uses reads mapped to 15 universal single-copy gene families (with ability to accurately recruit metagenomic reads to the correct species [31]) to estimate the abundance and coverage of bacterial species from a shotgun metagenome. We used the 132 *Nitrospira* species to build the database of universal-single-copy genes. Metagenomes were considered to contain *Nitrospira* spp. if at least eight reads mapped against the *Nitrospira* species present in our database of 15 universal single-copy genes. For richness calculation, the number of reads in each metagenome was normalized to 5 million and a *Nitrospira* species was considered to be present in a metagenome if at least an average (from 3 independent normalisations) of eight reads were mapped (as that would be above 50% of the 15 universal-single-copy genes used by MIDAS). For the co-occurrence analysis, only *Nitrospira* species with at least one read per million (RPM) mapped on at least six metagenomes (above 1% of the metagenomes where *Nitrospira* spp. were detected) were considered.

**Comparative genome analysis**

Two hundred five *Nitrospira* MAGs were included in the comparative genomic analysis. Gene calling was performed using Prodigal v.2.63 [72]. Annotation was conducted in RAST [73] and protein functional assignments of genes of interest were confirmed using blastp [74]. Pangenomic analysis was executed using the meta-pangenomic workflow of Anvi’o (v. 7.1) [30] with default parameters with the exception –maxbit = 0.3 (as described in Palomo et al. [7]). Briefly, blastp was used to calculate similarities between all proteins in each genome. Weak matches between two protein sequences were eliminated using maxbit heuristic [75]. Finally, the Markov Cluster Algorithm [76] was used to generate gene clusters (GCs) from protein similarity search results. GCs were considered part of the core *Nitrospira* genome when present in at least 80% of the genomes. GCs were considered enriched in comammox *Nitrospira* when present in more than 60% of the comammox genomes (at least 40% in each clade) and absent in more than 90% of the non comammox genomes. GCs were classified as clade-specific (clade A, clade B, clade A1, and clade A2) if present in at least 55% of the clade-type genomes and absent in the rest of the *Nitrospira* genomes (Table S3).

**Phylogenetic analysis and gene synteny**

Phylogenetic analyses of *Nitrospira* genomes were conducted with the GTDB-Tk v.0.1.3 tool [77] using the de novo workflow with a set of 120 translated universal single-copy genes and the genome taxonomy database (GTDB) [78]. Concatenated alignments were used to construct a maximum likelihood tree using RAxML v. 8.2.11 [79] with 200 rapid bootstraps (determined using the autoMRE option) and the LG likelihood model of amino acid substitution with gamma distributed rates and fraction of invariant sites (-m PROTGAMMALGF; all substitution models were selected using ProtTest v. 3.4.2 [80]). The tree was rooted using two *Leptospirillum* species as outgroup. The online web tool from the Interactive Tree of Life (iTol) [81] was used for visualization. Predicted AmoA and HaoA amino-acid sequences were independently aligned with reference sequences using MUSCLE [82]. These alignments were used to construct maximum likelihood trees using RAxML v. 8.2.11 with 900 and 300 rapid bootstraps, respectively (determined using the autoMRE option). For AmoA, the tree was built using the LG model plus gamma distribution of rates across sites considering the amino acid frequencies of the data set evolution (-m PROTGAMMALGF), while for HaoA the tree was constructed using LG model with an estimation of invariable sites and gamma distribution (-m PROTGAMMALGF). Both trees were rooted using beta- and gammaproteobacterial AOB species as outgroup. The rooted trees were visualized using the online web tool from the Interactive Tree of Life (iTol) [81].

Gene arrangement of ammonia oxidation related genes was visualized using the R package genoPlotR [83].

**Statistical analyses**

All statistical tests were performed using R v3.4.4 [84]. Due to the compositional nature of sequencing data [85], for all statistical analyses, species abundances were analyzed as follows: zeros were replaced with an estimate value using the count zero multiplicative approach with the zCompositions R package [86], and data were further centered log-ratio transformed. Statistical significance of the mean richness and abundance of *Nitrospira* species in the different habitats were assessed using Kruskal–Wallis ANOVA followed by Dunn’s test with the Holm–Bonferroni correction. Statistical significance of the mean richness and abundance between canonical and comammox *Nitrospira* species were evaluated using two-sided Mann-Whitney–Wilcoxon test.

Principal components analysis (PCA) was performed in R package factoextra [87]. Proportionality between abundances of the species across the 598 metagenomes were calculated using the propr R package [88] (with the options metric = “rho”, iivar = “clr”), which takes the compositional nature of the data into account, and visualized using the corrplot R package [89]. A Mantel test was used...
to investigate correlations between a matrix containing habitat similarity between species (as measured by the correlation between the abundances of *Nitrospira* species across samples) and a matrix of genome similarity (as measured by average nucleotide identity or sequence similarity of a set of 120 translated universal single-copy genes between pair of genomes) or amino acid sequence similarities of key nitrification enzymes (AmoA, HaoA, and NxrB). The significance of the Mantel statistic was obtained after 99,999 permutations.

* Nitrospira* community dissimilarities were calculated using the Jaccard index. The correlation between the *Nitrospira* community dissimilarities and geographic distances was calculated using the Mantel test (significance obtained after 99,999 permutations).

### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-022-01411-y.

#### Additional file 1: Figure S1. Pairwise average amino acid identity (AAI) of pairs of *Nitrospira* species. The AAI was clustered using average linkage hierarchical clustering based on pairwise Euclidean distances. Colour of the genome's name indicates the *Nitrospira* type (see Fig. 4). Figure S2. Frequency of *Nitrospira* species occurrence in the metagenomes (n = 598) where at least one *Nitrospira* species was detected. Black, grey and white colors denote 0.5, 5, and 15 RPM of each species in each metagenome, respectively. Figure S3. Relationship between the community similarity and the geographic distance for seven different habitats. The dissimilarities between pairs of communities are calculated using the Jaccard index from a presence/absence matrix of *Nitrospira* genomes; the value 0 means that the two communities are the same. The Mantel test was used to test the strength and significance of correlations (r denotes the Mantel statistic r). Blue line shows the linear regression with shadowed region indicating 95% confidence intervals for the slope. The table shows the correlation (Mantel statistic r) between the community similarity and the geographic distance when all samples were analysed and when samples within short distances were excluded. Figure S4. Heat map analysis of *Nitrospira* species abundances across 598 metagenomes. Dendrograms are built based on Euclidean distance. Rows represent individual metagenomes and columns represent unique *Nitrospira* species. Colour intensity indicates center-log ratio transformed abundance. Figure S5. Correlogram showing the proportionality (p) between the abundance of pairs of *Nitrospira* species across 598 metagenomes. Colour indicates whether the proportionality is positive (purple) or negative (brown). Size and darkness of the circles indicate the strength of the proportionality, with stronger proportionalities being larger and darker than weaker ones. A cut-off |p| > 0.15 was chosen as it resulted in FDR < 0.001. Colour of the *Nitrospira* species indicates the type (see Fig. 4). Figure S6. Unique shared synteny between clade A2 and some of the clade B genomes in the hydroxylamine reductase genomic region (yellow arrows; duplicated NirK (CuNIR) in wine colour). Schematic of the ammonia oxidation pathway genomic region in *comamox* *Nitrospira* clade A1, clade A2, and clade B genomes. Functions of the encoded proteins are represented by colour. Single diagonal line designates a break due to contig fragmentation. Arrows represent genes, arrow direction denotes the orientation of the coding strand, and arrow lengths are proportional to the gene lengths. Figure S7. The implemented workflow for the MAGs recovery from 12 Danish groundwater-fed rapid sand filters. The final genome quality improvement performed with MeDuSa was only applied on the 18 *Nitrospira* MAGs. The numbers in the binning algorithm boxes indicate the minimum contig size considered for the binning step. For MetaBAT, SupB20 and SpPb820 indicate “super-specific” and “very-sensitive” modes, respectively. Figure S8. Genetic discontinuity observed using 205 *Nitrospira* genomes. The scatter plot shows the pairwise MASH-based ANI (y axis) and ANIn (x axis) values among *Nitrospira* genomes. Only values in the 90–100% range are showed. Histogram plots show the distribution of MASH-based ANI and ANIn values among the *Nitrospira* genomes. A genetic discontinuum among the *Nitrospira* genomes is identified at around 96% for both metrics. Supplementary Table S1. List of publicly available metagenomes used to retrieve metagenome-assembled *Nitrospira* genomes. (provided as an Excel table). Supplementary Table S2. Characteristics of *Nitrospira* genomes. (provided as an Excel table). Supplementary Table S3. List of gene clusters of the *Nitrospira* pangenome. (provided as an Excel table). Supplementary Table S4. Characteristics of metagenomes. (provided as an Excel table). Supplementary Table S5. Richness and abundance of 132 *Nitrospira* species across habitats.

### Acknowledgements

The authors acknowledge the contributors to the Sequence Read Achieve and ViG-RAST for making their data publicly available. This study would not have been possible without this open sharing of data. The authors would like to acknowledge George Kuwarteng Armoako (HOFOR), Stig Eskildsen (Forsyning Ballerup), Steffen Schulz and Jørgen Bentsen (Nordvand), Torben Snehaj Weng (Fors A/S), Jens Kjølby Larsen (Forsyningen), and Dr. Marta Kinnunen for their support in sample and data collection, as well as Dr. Florian B. Wagner and Dr. Vaibhav Diwan for providing samples from some of the waterworks. We are grateful to Neillihon Bicen and Marlene Danner Dalgaard at DTU MultiAssay Core (DMAC) Institute for Health technology for supporting library preparation and sequencing. We would also like to thank Dr. Jan-Ulrich Kröft and Dr. Andreas Schramm for their useful feedback on this study.

### Authors’ contributions

A.P, A.D, and B.F.S conceived the study. A.P performed the bioinformatic analyses. A.P and A.D led interpretation of the results supported by B.F.S and A.G.P. A.P drafted the manuscript with input from A.D and B.F.S. All authors contributed to manuscript revision and approved the final version of the manuscript.

### Funding

This research was supported by a research grant (13391, Expa-N) from VILLUM FONDEN.

### Availability of data and materials

All raw sequence data and genome sequences retrieved from the Danish rapid sand filters have been deposited at NCBI under the project PRJNA384587. The rest of the retrieved draft genomes are available on figshare (https://doi.org/10.6084/m9.figshare.7999448). The file containing the gene clusters sequences of the *Nitrospira* pangenome is available on figshare (https://doi.org/10.6084/m9.figshare.7998641). The data that support the plots within this paper are available from corresponding author upon reasonable request.

### Declarations

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1Department of Environmental and Resource Engineering, Technical University of Denmark, Kgs Lyngby, Denmark. 2School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China. 3Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Copenhagen, Denmark.

Received: 12 August 2022 Accepted: 3 November 2022 Published online: 30 November 2022
References

1. Cassman KG, Dobermann A, Walters DT. Agroecosystems, nitrogen-use efficiency, and nitrogen management. AMBIO A J Hum Environ. 2002;31:132–40.

2. Okabe S, Aoi Y, Satoh H, Suwa Y. Nitritification in wastewater treatment. In: Ward B, Arp D, Klots M (eds). Nitritification. 2011. American Society of Microbiology, Washington, DC, pp 405–433.

3. Simon J, Klots MG. Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. Biochim Biophys Acta - Bioenerg. 2013;1827:114–35.

4. Daims H, Lebedeva EV, Pijevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by Nitrospira bacteria. Nature. 2015;528:504–9.

5. van Kessel MAH, Speth DR, Albertsen M, Nielsen PH, Op den Camp HJM, Kartal B, et al. Complete nitrification by a single microorganism. Nature. 2015;528:555–9.

6. Daims H, Lucke S, Wagner M. A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. Trends Microbiol. 2016;24:699–712.

7. Palomo A, Pedersen AG, Fowler SJ, Dechesne A, Sichentz-Ponten T, Smets BF. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox Nitrosira. ISME J. 2018;12:1779–93.

8. Costa E, Pérez J, Keff J-U. Why is metabolic labour divided in nitrification? Trends Microbiol. 2006;14:213–9.

9. Lawson CE, Lücker S. Complete ammonia oxidation: an important control on nitrification in engineered ecosystems? Curr Opin Biotechnol. 2018;50:158–65.

10. Pester M, Maixner F, Berry D, Ratté T, Koch H, Lucke S, et al. Nitr encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing Nitrospira. Environ Microbiol. 2014;16:3055–71.

11. Hovanec TA, Taylor LT, Blakis A, Delong EF. Nitrospira-like bacteria associated with nitrite oxidation in freshwater aquaria. Appl Environ Microbiol. 1998;64:258–64.

12. Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U. Nitrosira marina gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. Arch Microbiol. 1986;144:1–7.

13. Schwab VB, Herrmann M, Roth-V-N, Gleixner G, Lehmann R, Pohner N, et al. Functional diversity of microbial communities in pristine aquifers inferred by PLFA - and sequencing - based approaches. Biogeoosciences. 2016;14:2697–714.

14. White CP, Debyr RW, Lytle D, a. Microbial survey of a full-scale, biologically active filter for treatment of drinking water. Appl Environ Microbiol. 2012;78:6390–4.

15. Gülay A, Musovic S, Albrechtson H-I, Al-Soud WA, Sørensen SI, Smets BF. Ecological patterns, diversity and core taxa of microbial communities in groundwater-fed rapid gravity filters. ISME J. 2016;10:2209–22.

16. Pijevac P, Schauberger C, Poghosyan L, Herbold CW, van Kessel MAH, Daebeler A, et al. AmoA-targeted polymerase chain reaction primers for the specific detection and quantification of comammox nitrospira in the environment. Front Microbiol. 2017;8:1508.

17. Sun D, Tang X, Zhao M, Zhang Z, Hou L, Liu M, et al. Distribution and diversity of comammox nitrospira in coastal wetlands of China. Front Microbiol. 2020;11:589268.

18. Zhang M, Li Z, Hägglomb MM, Young L, He Z, Li F, et al. Characterization of nitrate-dependent As(III)-oxidizing communities in arsenic-contaminated soil and investigation of their metabolic potentials by the combination of DNA-stable isolate probing and metagenomics. Environ Sci Technol. 2020;54:7366–77.

19. Hu J, Zhao Y, Yao X, Wang J, Zhang P, Xi G, et al. Dominance of comammox Nitrospira in soil nitrification. Sci Total Environ. 2021;780:146558.

20. Bartelme RF, McCellan SL, Newton RJ. Freshwater recirculating aquaculture system operations drive biofilter bacterial community shifts around a three-vessel system. Water Res. 1998;32:769–73.

21. Zheng M, Wang M, Zhao Z, Huang P, He S, Liu Y, et al. Transcriptional activity and diversity of comammox bacteria in a previously overlooked ammonia oxidizing protoxynote in full-scale wastewater treatment plants. Sci Total Environ. 2019;656:717–22.

22. Roots P, Wang Y, Rosenthal AF, Griffin JS, Sabbia F, Petrovich M, et al. Comammox Nitrosospira: An advanced analysis and visualization platform for omics data. PeerJ. 2015;3:e1319.

23. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. Ye- round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl Environ Microbiol. 2017;83:1646–17.

24. Osburn ED, Barrett JE. Abundance and functional importance of complete ammonia-oxidizing bacteria (comammox) versus canonical nitrifiers in temperate forest soils. Soil Biol Biochem. 2020;145:107801.

25. Xia F, Wang J-G, Zhu T, Zou B, Rhee S-K, Quan Z-X. Ubiquity and Diversity of Complete Ammonia Oxidizers (Comammox). Appl Environ Microbiol. 2018;84:1–14.

26. Shi Y, Jiang Y, Wang S, Wang X, Zhu G. Biogeographic distribution of comammox bacteria in diverse terrestrial habitats. Sci Total Environ. 2020;717:137257.

27. Palomo A, Jane Fowler S, Gülay A, Rasmussen S, Sichentz-Ponten T, Smets BF. Metagenomic analysis of rapid gravity sand filter microbial communities suggests novel physiology of Nitrosira spp. ISME J. 2016;10:2569–81.

28. Wang Y, Ma L, Mao Y, Jiang X, Xia Y, Ku K, et al. Comammox in drinking water systems. Water Res. 2017;116:332–41.

29. Liu S, Wang H, Chen L, Wang J, Zheng M, Liu S, et al. Comammox Nitrosira within the Yangtze River continuum: community, biogeography, and ecological drivers. ISME J. 2020;14:2488–504.

30. Eren AM, Eisen OC, Quince C, Vines JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for omics data. PeerJ. 2015;3:e1319.

31. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016;26:1612–25.

32. Delmont TO, Quince C, Shailer A, Eisen OC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol. 2018;3:804–13.

33. Off S, Alavi M, Speck E. Enrichment and physiological characterization of a novel nitrite-oxidizing bacterium obtained from a marine sponge. Appl Environ Microbiol. 2010;76:4640–6.

34. Lebedeva EV, Alavi M, Maixner F, Jozsa P-G, Daims H, Speck E. Physiological and phylogenetic characterization of a novel lithoautotrophic nitrite-oxidizing bacterium, Candidatus Nitrosospira bocciarum. Int J Syst Evol Microbiol. 2008;58:242–50.

35. Lebedeva EV, Off S, Zumbelügel S, Kruse M, Shaghzina A, Lucke S, et al. Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring. FEBS Microbiol Ecol. 2011;75:195–204.

36. Lautenschlager K, Hwang C, Ling F, Liu W-T, Boon N, Köster O, et al. Functional diversity of microbial communities in pristine aquifers inferred by PLFA - and sequencing - based approaches. Biogeoosciences. 2016;14:2697–714.

37. Kits KD, Sedlacek J, Lebedeva EV, Han P, Bulava A, Pijevac P, et al. Kinetic analysis of a complete nitrite reductase from novel ammonia oxidizers responding to incompatible nitrification activities in three-vessel system. Water Res. 1998;32:769–73.

38. Zheng M, Wang M, Zhao Z, Zhou N, He S, Liu Y, et al. Transcriptional activity and diversity of comammox bacteria in a previously overlooked ammonia oxidizing protoxynote in full-scale wastewater treatment plants. Sci Total Environ. 2019;656:717–22.

39. Roots P, Wang Y, Rosenthal AF, Griffin JS, Sabbia F, Petrovich M, et al. Comammox Nitrosospira: An advanced analysis and visualization platform for omics data. PeerJ. 2015;3:e1319.

40. Cotto I, Dai Z, Hlu L, Anderson CL, Vilarid KI, Jia Z, et al. Long solids retention times and attached growth phase favor prevalence of comammox bacteria in nitrogen removal systems. Water Res. 2020;169:115288.

41. Jiang G, Xia F, Zhu T, Wang D, Quan Z. Distribution of comammox and canonical ammonia-oxidizing bacteria in tidal flat sediments of the Yangtze River estuary at different depths over four seasons. J Appl Microbiol. 2019;127:533–43.
