Complete nitrification by a single microorganism

Maartje A. H. J. van Kessel1, Daan R. Speth1, Mads Albertsen2, Per H. Nielsen2, Huub J. M. Op den Camp1, Boran Kartal1,3, Mike S. M. Jetten1,4 & Sebastian Lücker1

Nitrification is a two-step process where ammonia is first oxidized to nitrite by ammonia-oxidizing bacteria and/or archaea, and subsequently to nitrate by nitrite-oxidizing bacteria. Already described by Winogradsky in 1890, this division of labour between the two functional groups is a generally accepted characteristic of the biogeochemical nitrogen cycle. Complete oxidation of ammonia to nitrate in one organism (complete ammonia oxidation; comammox) is energetically feasible, and it was postulated that this process could occur under conditions selecting for species with lower growth rates but higher growth yields than canonical ammonia-oxidizing microorganisms. Still, organisms catalysing this process have not yet been discovered. Here we report the enrichment and initial characterization of two Nitrospira species that encode all the enzymes necessary for ammonia oxidation via nitrite to nitrate in their genomes, and indeed completely oxidize ammonium to nitrate to conserve energy. Their ammonia monoxygenase (AMO) enzymes are phylogenetically distinct from currently identified AMOs, rendering recent acquisition by horizontal gene transfer from known ammonia-oxidizing microorganisms unlikely. We also found highly similar amoA sequences (encoding the AMO subunit A) in public sequence databases, which were apparently misclassified as methane monoxygenases. This recognition of a novel amoA sequence group will lead to an improved understanding of the environmental abundance and distribution of ammonia-oxidizing microorganisms. Furthermore, the discovery of the long-sought-after comammox process will change our perception of the nitrogen cycle.

Nitrification, the aerobic oxidation of ammonium to nitrate, is divided into two subsequent reactions: ammonium oxidation to nitrite (equation (1)) and nitrite oxidation to nitrate (equation (2)). These two reactions are catalysed by physiologically distinct clades of microorganisms.

\[
\begin{align*}
\text{NH}_4^+ + 1.5 \text{O}_2 & \rightarrow \text{NO}_2^- + \text{H}_2\text{O} + 2\text{H}^+ & (\Delta G^\circ = -274.7 \text{ kJ mol}^{-1}) \quad (1) \\
\text{NO}_2^- + 0.5 \text{O}_2 & \rightarrow \text{NO}_3^- & (\Delta G^\circ = -74.1 \text{ kJ mol}^{-1}) \quad (2) \\
\text{NH}_4^+ + 2\text{O}_2 & \rightarrow \text{NO}_3^- + \text{H}_2\text{O} + 2\text{H}^+ & (\Delta G^\circ = -348.9 \text{ kJ mol}^{-1}) \quad (3)
\end{align*}
\]

Even though the existence of a single microorganism capable of oxidizing ammonium to nitrate was not previously reported, it was proposed that such a microorganism could have a competitive advantage in biofilms and other microbial aggregates with low substrate concentrations.

In this study, to characterize the microorganisms responsible for nitrogen transformations in an ammonium-oxidizing biofilm, we sampled the anaerobic compartment of a trickling filter connected to a recirculation aquaculture system with an ammonium effluent of less than 100 μM. To enrich for the N-cycling community, a bioreactor was inoculated and supplied with low concentrations of ammonium, nitrite and nitrate under hypoxic conditions (≤3.1 μM O2). Within 12 months, we obtained a stable enrichment culture that efficiently removed ammonium and nitrite from the medium (Extended Data Fig. 1). The culture showed anaerobic ammonium-oxidizing (anammox) activity (Fig. 1a), and fluorescence in situ hybridization (FISH) revealed that anammox organisms of the genus Brocadia constituted approximately 45% of all FISH-detectable bacteria. Surprisingly, Nitrospira-like nitrite-oxidizing bacteria accounted for approximately 15% of the community and co-occurred with the Brocadia species in flocs (Fig. 2a). This tight clustering with anammox bacteria was unexpected as both microorganisms require nitrite for growth. Together with the presence of Nitrospira at very low oxygen concentrations, this indicated that there could be a functional link between these organisms.

To determine the function of Nitrospira in the community, we extracted and sequenced total DNA from the enrichment culture biomass. In total 4.95 gigabase pairs of trimmed metagenomic sequence were obtained and used for de novo assembly. By differential coverage and sequence composition-based binning it was possible to extract high-quality draft genomes of two Nitrospira species. The two strains had genomic pairwise average nucleotide identities (ANI) of 75% and thus clearly represented different species (Nitrospira sp.1 and sp.2, Extended Data Fig. 2 and Extended Data Table 1). Surprisingly, both genomes contained the full set of AMO and hydroxylamine dehydrogenase (HAO) genes for ammonia oxidation, in addition to the nitrite oxidoreductase (NXR) subunits necessary for nitrite oxidation in Nitrospira. In both species all these genes were localized on a single contiguous genomic fragment, along with general housekeeping genes that allowed reliable phylogenetic assignment. Consequently, these Nitrospira species had the genetic potential for the complete oxidation of ammonia to nitrate. No AMO of canonical ammonia-oxidizing bacteria or archaea could be detected in the trimmed metagenomic reads or by amoA-specific PCR on DNA extracted from reactor biomass, and no other indications for the presence of ammonia-oxidizing microorganisms were found in the metagenome or by FISH analyses. The AMO structural genes (amoCAB) of both Nitrospira species, along with the putative additional AMO subunits amoEDD21,11, formed one gene cluster with haoAB-cycAB (encoding HAO, the putative membrane anchor protein HaoB, electron transfer protein cytochrome c₅₅₄ and quinone-reducing cytochrome c₅₅₂, respectively)12 and showed highest similarities to their counterparts in betaproteobacterial ammonia-oxidizing bacteria (60% average amino acid identity to the Nitrosomonas europaea genes; Fig. 3 and Supplementary Table 1). The same genomic region also contained genes for copper and haem transport, cytochrome c biosynthesis, and iron storage. These accessory genes were highly conserved in ammonia-oxidizing bacteria but not in other Nitrospira13, indicating their involvement in AMO and HAO biosynthesis or activation. Nitrospira sp.1 encoded three discrete amoC genes, one of which was clustered with a second, almost identical copy of amoC (97.7% amino acid identity). Nitrospira sp.2 lacked the second amoA, but contained four additional amoC and a second haoA gene (Supplementary Table 1). Unlike other Nitrospira13, both species lacked enzymes for assimilatory nitrite reduction, indicating adaptation to ammonium-containing
habits. For ammonium uptake, they encoded low-affinity Rh-type transporters most closely related to Rh50 found in *Nitrosomonas europea*14, in contrast to most ammonia-oxidizing and nitrite-oxidizing bacteria that have the high-affinity AmoB-type proteins. Both species encoded ureases and the corresponding ABC transport systems, indicating that urea could be used as an alternative ammonium source. Interestingly, *Candidatus Nitrospira* inopinata, the moderately thermophilic ammonia-oxidizing *Nitrospira* described by Daims *et al.*15, encoded a similar set of AMO, HAO and urease proteins, and also lacked genes for assimilatory nitrite reduction. Unlike the two species described here, however, it contained a periplasmic cytochrome c nitrite reductase (NrfA) that could allow it to conserve energy by dissipatory nitrite reduction to ammonium (DNRA), but might also provide ammonium for assimilation. The evolutionary divergence of these organisms was also reflected in the low ANI values of 70.3–71.6% between *Candidatus* N. inopinata and the two species described here. Concerning their genetic repertoire for nitrite oxidation, sp.2 had four almost identical (>99% amino acid identity) NXR alpha and beta (NxrAB) subunits. Sp.1 had two nxrAB copies encoding identical NxrB subunits, but NxrA subunits with amino acid identities of 89.6%, which were separated into distinct clusters in phylogenetic analyses. One homologue branched with sequences from *Nitrospira moscowiensis*, while the other formed a novel sequence cluster together with the sequences from sp.2 (Extended Data Fig. 3).

To ascertain that ammonia oxidation occurred under hypoxic conditions in the enrichment culture, we supplied the bioreactor with 15N-labelled ammonium. While the anammox bacteria consumed 15NH4⁺ and converted it into 29N2, a steady increase of 30N2 was also observed (Fig. 1a). This formation of 30N2 could only be explained by the production of 15N-labelled nitrite derived through aerobic ammonium oxidation. As metagenomic analyses confirmed that the *Nitrospira* species were the only organism in the enrichment harbouring AMO and HAO, this clearly showed that they were able to perform this reaction even at O2 concentrations lower than 3.1 μM. To unambiguously link this activity to *Nitrospira*, we visualized the AMO protein in situ using batch incubations with reactor biomass and FTCP (fluorescein thiocarbamoylpropargylamine), a fluorescently labelled acetylene analogue that acts as suicide substrate for AMO16 and covalently binds to the enzyme17. When counterstained with *Nitrospira*-specific FISH probes, including a newly designed probe specifically targeting anammox bacteria (Amx820, red, resulting in magenta), images in b and c are representative of two individual experiments, with three (b) or two (c) technical replicates each. Scale bars in all panels represent 10 μm.

**Figure 1** | Ammonium oxidation by the enrichment culture. a, 29N2 (open circles) and 30N2 (filled circles) production from 15NH4⁺ by the enrichment culture. b, Ammonium (diamonds) oxidation to nitrate (squares) in aerobic batch incubations in the absence (filled symbols) and presence (open symbols) of ATU. Nitrite concentrations were below the detection limit (<5 μM) at all time points. c, Nitrite (triangles) oxidation to nitrate (squares) in aerobic batch incubations. In b and c, total nitrogen balances are indicated (dashed lines). Symbols in all plots represent averages of three individual experiments. Ammonium concentrations were determined in single measurements, other compounds in triplicate. Error bars represent standard deviations of three biological replicates.

**Figure 2** | In situ detection of *Nitrospira* and their ammonia-oxidizing capacity. a, Co-aggregation of *Nitrospira* and Brocadia in the enrichment. Cells are stained by FISH with probes for all bacteria (EUB338mix, blue), and specific for *Nitrospira* (Ntspa712, green, resulting in cyan) and anammox bacteria (Amx820, red, resulting in magenta). b, AMO labelling by FTCP (green). *Nitrospira* was counterstained by FISH (probes Ntspa662 (blue) and Ntspa476 (red), resulting in white). c, Ammonium-dependent CO2 fixation by *Nitrospira* shown by FISH–MAR. Silver grain deposition (black) above cell clusters indicates 14CO2 incorporation. *Nitrospira* was stained by FISH (probes Ntspa476 (red) and Ntspa662 (blue), resulting in magenta). Images in b and c are representative of two individual experiments, with three (b) or two (c) technical replicates each. Scale bars in all panels represent 10 μm.
the presence of the ammonia-oxidizing enzyme at the single-cell level (Fig. 2b and Extended Data Fig. 5).

Batch incubations were performed at ambient oxygen concentrations to determine conversion rates of ammonium and nitrite, the level of inhibition by allylthiourea (ATU; a potent inhibitor of bacterial ammonia oxidation\textsuperscript{18,19}), and the use of urea as ammonium source for nitrification. Flocs were mechanically disrupted to ensure complete exposure of the biomass to oxygen, which inhibits the amnonox and denitrification processes\textsuperscript{20,21}. This inhibition was confirmed by the lack of labelled N\textsubscript{2} formation in incubations with 15NH\textsubscript{4}\textsuperscript{+} and nitrite (23 ± 4.7 μM). In these incubations (Fig. 1 and Extended Data Fig. 6), the culture oxidized ammonium (6.0 ± 1.0 μM NH\textsubscript{4}\textsuperscript{+}) and nitrite (23 ± 4.7 μM NO\textsubscript{2}\textsuperscript{−}) to nitrate. ATU selectively inhibited ammonia oxidation, but did not affect nitrite oxidation rates. Urea was converted to ammonium, which was subsequently oxidized to nitrate (7.8 ± 1.1 μM). Given these characteristics, it was suggested that these Nitrospira species were capable of using urea as source of ammonia to drive nitrification, as was also reported for some ammonia-oxidizing archaea\textsuperscript{22} and bacteria\textsuperscript{23}. This trait could enable them to thrive in environments like fertilized soils, wastewater treatment plants, and many aquatic systems where urea is often present at micromolar levels\textsuperscript{24}. However, it should be noted that the two Nitrospira spp. were not the only organisms in the enrichment culture that encoded ureases.

To investigate substrate-dependent inorganic carbon fixation as a proxy for energy conservation from ammonia and nitrite oxidation, we used FISH in combination with microautoradiography (FISH-MA\textsuperscript{R})\textsuperscript{25}. Aerobic incubations with mechanically disrupted flocs were performed in the presence of 500 μM ammonium, 500 μM ammonium with 100 μM ATU, or 500 μM nitrite. Nitrospira incorporated carbon from 14C-labelled bicarbonate in the presence of either ammonium or nitrite, and ammonia-dependent carbon fixation was strongly inhibited by the addition of ATU (Fig. 2c and Extended Data Fig. 7). Only flocs containing Nitrospira were labelled during all incubations, suggesting that these were the only chemolithoautotrophic nitrifying organisms present in the culture and indeed could conserve energy from the oxidation of ammonia and nitrite.

In 16S rRNA-based phylogenetic analyses, the two ammonia-oxidizing Nitrospira species from our enrichment culture formed two separate lineages within one strongly supported sequence cluster affiliated with Nitrospira sublineage II\textsuperscript{26} (Extended Data Fig. 4). They both grouped with highly similar sequences (>99% nucleotide identity) from a diverse range of habitats, including soil, groundwater, recirculation aquaculture systems, wastewater treatment plants and drinking water distribution systems. The formation of distinct clusters containing sp.1 and sp.2 indicated that the last common ancestor encoded genes for complete nitrification and that this pathway might be conserved in most organisms affiliated with this sequence group.

To explore the environmental relevance of these Nitrospira, we searched the NCBI nr database\textsuperscript{27} for closely related amoA genes. Surprisingly, we found the AmoA proteins of the two Nitrospira species to be phylogenetically divergent from the described bacterial AmoA sequences. Nitrospira sp.2 AmoA was 97–98% identical to the so-called ‘unusual’ methane monoxygenase (PMO) proteins of Crenothrix polyspora\textsuperscript{28}. The two AmoA copies from Nitrospira sp.1 had lower similarities to Crenothrix PmoA (90–91% identity), but also affiliated with this group (Fig. 4). Sequences within this group cannot be amplified by standard amoA primers, but only by pmoA primers when used at reduced stringency\textsuperscript{29}. Therefore, the public databases only contain few closely related sequences, which were mainly derived from habitats studied for their bacterial methane-oxidizing communities. Highly similar sequences derived from wastewater treatment plants and drinking water systems, however, indicated occurrence of ammonia-oxidizing Nitrospira in a range of engineered and natural environments.

We furthermore screened all publicly available shotgun data sets on MG-RAST\textsuperscript{30}. Indeed, 168 metagenomes (out of 6,255) and 28 metatranscriptomes (out of 1,051) contained at least two reads affiliated with this amoA group, yielding a total of 3,727 reads that were obtained mainly from soil, sediments and wastewater treatment plants (Extended Data Table 3). Thus, our results showed that the Crenothrix sequence group consists of so far unrecognized AMO sequences overlooked in nitrification studies based on amoA gene detection. Based on these findings, it is highly likely that the PCR-based determination of the Crenothrix pmoA gene from an environmental sample\textsuperscript{26} was erroneous, and this cluster only contains genes encoding AMOs. Nevertheless, with the currently available information it cannot be excluded that certain Crenothrix species retained an amoA gene through lateral gene transfer and use the encoded protein as a surrogate PMO.

In conclusion, here we demonstrated the existence of complete nitrification in a single organism (comammox) and identified two Nitrospira species capable of catalysing this process (equation (3)). In 16S rRNA or amoA/pmoA-based studies these organisms would have been classified as nitrite-oxidizing or methane-oxidizing bacteria, respectively. Hence, our results show that a whole group of ammonia-oxidizing organisms was previously overlooked. Our findings furthermore disprove the long-held assumption that nitrification (ammonia oxidation via nitrite to nitrate) is catalysed by two distinct functional groups, thus redefining a key process of the biogeochemical nitrogen cycle.

Based on their phylogeny, differences in genome content, and separation in different phylogenetic groups in 16S rRNA-based analyses,
we propose tentative names for both *Nitrosira* species present in our enrichment: *Candidateatus* Nitrosira nitrosa (etymology: L. fem. adj. *nitrosa*, nitrous; the nitrite and nitrate forming *Nitrosira*) for sp.1 and *Candidateatus* Nitrosira nitrificans (N.L. part. adj. *nitrificans*, nitrifying; the nitrifying *Nitrosira*) for sp.2. Both species are chemolithoautotrophic and fully oxidize ammonia via nitrite to nitrate.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to Methods, along with any additional Extended Data display items and Online Content are available in the online version of the paper; references unique to Online Content are available in the online version of the paper.

**Received 10 August; accepted 18 November 2015.** Published online 26 November 2015.

---

1. Winogradsky, S. Recherches sur les organismes de la nitrification. Ann. Inst. Pasteur (Paris) 4, 213–231 (1890).
2. Vlaeminck, S. E., Hay, A. G., Maingien, L. & Verstraete, W. In quest of the nitrogen oxidizing prokaryotes of the early Earth. Environ. Microbiol. 13, 283–295 (2011).
3. Costa, E., Pérez, J. & Kreft, J. U. Why is metabolic labour divided in nitrification? Trends Microbiol. 14, 213–219 (2006).
4. Grab, R., Avnimelech, Y., Deforot, T., Bossier, P. & Verstraete, W. Nitrogen removal techniques in aquaculture for a sustainable production. Aquaculture 270, 1–14 (2007).
5. Albertsen, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nature Biotechnol. 31, 533–538 (2013).
6. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl Acad. Sci. USA 106, 19126–19131 (2009).
7. Lücker, S. et al. *Nitrosira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. Proc. Natl Acad. Sci. USA 107, 13479–13484 (2010).
8. Rothhauwe, J. H., Witzel, P. K. & Lesack, W. The ammonia monoxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl. Environ. Microbiol. 63, 4704–4712 (1997).
9. Könneke, M. et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437, 543–546 (2005).
10. El Sheik, A. F., Poret-Peterson, A. T. & Klotz, M. G. Characterization of two new genes, amoR and amoD, in the amo operon of the marine ammonia oxidizer Nitrooccus oceanis ATCC 19707. Appl. Environ. Microbiol. 74, 312–318 (2008).
11. Berube, P. M. & Stahl, D. A. The divergent AmoC3 subunit of ammonia monoxygenase functions as part of a stress response system in *Nitrosomonas europaea*. J. Bacteriol. 194, 3448–3456 (2012).
12. Klotz, M. G. & Stein, L. Y. Nitrifier genomics and evolution of the nitrogen cycle. FEMS Microbiol. Lett. 278, 145–156 (2008).
13. Koch, H. et al. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrosira*. Proc. Natl Acad. Sci. USA 112, 11371–11376 (2015).
14. Lupo, D. et al. The 1.3-A resolution structure of Nitrosomonas europaea Rf50 and mechanistic implications for NH3 transport by Rhus family proteins. Proc. Natl Acad. Sci. USA 104, 19303–19308 (2007).
15. Daims, H. et al. Complete nitrification by Nitrosirabacteria. Nature http://dx.doi.org/10.1038/nature14641 (2015).
16. McTeish, H., Fuchs, J. A. & Hooper, A. B. Sequence of the gene coding for ammonia monoxygenase in *Nitrosomomas europaea*. J. Bacteriol. 175, 2436–2444 (1993).
17. Hyman, M. R. & Arp, D. J. 14C2H2- and 14CO2-labeling studies of the de novo synthesis of polyaldehydes by *Nitrosomomas europaea* during recovery from acetylene and light inactivation of ammonia monoxygenase. J. Biol. Chem. 267, 1534–1545 (1992).
18. Taylor, A. E. et al. Use of aliphatic n-alkenes to discriminate soil nitrification activities of ammonia-oxidizing thaumarchaea and bacteria. Appl. Environ. Microbiol. 79, 6544–6551 (2013).
19. Ginevit, P., Audic, J.-M., Urbain, V. & Block, J.-C. Estimation of nitrifying bacterial activities by measuring oxygen uptake in the presence of the metabolic inhibitors allythiocrylcarboxylic acid. Appl. Environ. Microbiol. 64, 2266–2268 (1998).
20. Zufrit, W. G. Cell biology and molecular basis of denitrification. Microbiol. Mol. Biol. Rev. 61, 533–616 (1997).
21. Stross, M., Kuenen, J. G. & Jetten, M. S. Key physiology of anaerobic ammonium oxidation. Appl. Environ. Microbiol. 65, 3248–3250 (1999).
22. Alfonso-Sáez, L. Role for urea in nitrification by polar marine Archaea. *Proc. Natl Acad. Sci. USA* 109, 17989–17994 (2012).
23. Burton, S. A. Q. & Prosser, J. L. Autotrophic ammonia oxidation at low pH through urea hydrolysis. *Appl. Environ. Microbiol.* 67, 2952–2957 (2001).
24. Solomon, C., Collier, J., Berg, G. & Gilbert, P. Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. Aquat. Microb. Ecol. 59, 67–88 (2010).
25. Wagner, M., Nielsen, P. H., Loy, A., Nielsen, J. L. & Damgaard, H. Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays. Curr. Opin. Microbiol. 17, 83–91 (2006).
26. Daims, H., Nielsen, J. L., Nielsen, P. H., Schleifer, K. H. & Wagner, M. *In situ* characterization of Nitrosis-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl. Environ. Microbiol.* 67, 5273–5284 (2001).
27. Johnson, M. et al. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9 (2008).
28. Stoecker, K. et al. Cohn’s Crenothrix is a filamentous methane oxidizer with an unusual methane monoxygenase. *Proc. Natl Acad. Sci. USA* 103, 2363–2367 (2006).
29. Luesken, F. A. et al. Diversity and enrichment of nitrite-dependent anaerobic methane oxidizing bacteria from wastewater sludge. *Appl. Microbiol. Biotechnol.* 92, 845–854 (2011).
30. Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D. & Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb. Protoc. http://dx.doi.org/10.1101/pdb.prot5368 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank K. Stultiens, T. van Alen, J. Frank, P. Klaren, L. Pierson and L. Claessens-Joosten for technical assistance, T. Spanings for biofilter maintenance and C. Herbold for the ANI analysis. We are grateful for the use of the confocal microscope from the Microscopic Imaging Centre (MIC, Radboud UMC, Nijmegen) and would like to thank H. Croes and M. Willemse for technical assistance. The LABGeM team and the National Infrastructure “France Genomique” are acknowledged for support within the MicroScope annotation platform. We are thankful to C. Dupont, A. Santoro and M. Saito for consenting to our use of the Nitrospira marina nxrA sequences, which were produced by the US Department of Energy Joint Genome Institute. M.A.H.J.v.K was supported by the Technology Foundation STW (grant 13146), D.R.S. by the BE-Basic Foundation (grant fs7-002), M.A. and P.H.N. by the Danish Council for Independent Research (OFF 4005-00369), M.S.M.J. by the European Research Council (ERC Advanced Grant projects anammox 232937 and Eco_MoM 339880) and the Dutch Ministry of Education, Culture and Science (Gravitation grant SIAM 024002002), B.K. and S.L. by the Netherlands Organization for Scientific Research (NWO VENI grants 863.11.003 and 863.14.019, respectively). The Radboud Excellence Initiative is acknowledged for support to S.L.

Author Contributions M.A.H.J.v.K and S.L. executed experiments and analysed data. D.R.S. and M.A. contributed to metagenomic data analyses. M.A. and P.H.N. performed sequencing, assembly and binning, M.A.H.J.v.K., H.J.M.O.d.C., B.K., M.S.M.J. and S.L. planned research. M.A.H.J.v.K., B.K. and S.L. wrote the paper. All authors discussed results and commented on the manuscript.

Author Information Metagenomic data is available in the European Nucleotide Archive (ENA) under accession numbers CZQA01000001–CZQA01000015 and CZPZ01000001–CZPZ01000036. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L. (s.luecker@science.ru.nl).
METHODS

No statistical methods were used to predetermine sample size, the experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment

Enrichment and cultivation. A bioreactor was inoculated with biomass from a recirculation aquaculture system biofilter (3.5 L, obtained from the anoxic part of the trickling filter compartment) connected to an aquaculture system. The system accommodated common carp (Cyprinus carpio, approximately 3.5 kg total weight) at a final volume of 900 L. The bioreactor (Applikon Biotechnology BV, Schiedam, The Netherlands) consisted of stainless steel and glass, had a 7 L working volume, was equipped with pH and dissolved oxygen sensors (Applikon Dependable Instruments BV Applisens, Schiedam, The Netherlands) and connected to an ADI1030 biocontroller (Applikon Biotechnology BV, Schiedam, The Netherlands). It was operated as a sequencing batch reactor (SBR) with 12 or 24 h cycles. In the first 5 months, the reactor was operated with a 24 h cycle that consisted of 23 h 15 min filling, 15 min settling (no stirring) and 30 min removal of the supernatant. Afterwards, in 12 h cycles, each filling cycle consisted of 11 h 15 min, followed by 15 min settling and 30 min removal of the supernatant. During every filling period, the reactor was supplied with 600 ml of medium (0.93 ml min⁻¹). The reactor and the medium were flushed constantly with Ar/CO₂ (95%/5% v/v, 10 ml min⁻¹). The temperature was kept at 23 ± 1°C with a heating blanket and pH was maintained at 6.99 ± 0.1 using a 1 M KHCO₃ solution. The reactor was stirred at 200 r.p.m. Medium was prepared using aquaculture water taken from the recirculation aquaculture system biofilter. This water contained 300–1,848 μM NO₃⁻, 0–29 μM NO₂⁻ and 0–75 μM NH₄⁺. The water was filter-sterilized (polysulfone filter HF805, Fresenius Medical Care, Bad Homburg, Germany) and supplemented with 100–500 μM NH₄⁺, 100–450 μM NO₃⁻ and 500 μM NO₂⁻.

DNA extraction and genome sequencing. DNA was extracted using the PowerSoil DNA isolation kit (MoBio, Carlsbad, CA) or a CTAB-based extraction method13. 1 μg of DNA was used to prepare paired-end sequencing libraries using the TrueSeq PCR-free kits (Illumina, San Diego, CA, USA) following the manufactures recommendation except that the 550 bp protocol was used with 1 μg of input DNA. Mate-pair libraries were prepared using the Nextera Mate-pair kit (Illumina) using the gel-free approach. The prepared libraries were sequenced using an Illumina MiSeq with MiSeq Reagent Kit v3 (2 × 301 bp; Illumina).

Bioinformatics. Data generation and binning of metagenome scaffolds to individual genome bins was conducted as described in the metagenome workflow5 which builds on the multi-metagenome principles5. Paired-end Illumina reads in FASTQ format were imported to CLC Genomics Workbench v.8.0 (CLCBio, Aarhus, Denmark) and trimmed using a minimum phred score of 20, a minimum length of 50 bp, allowing no ambiguous nucleotides and trimming off Illumina sequencing adaptors. Mate-pair reads in FASTQ format were trimmed using NextClip13 and only reads in class A were used for assembly. Passing reads were co-assembled using CLCs de novo assembly algorithm, using a kmer of 63 and a minimum scaffold length of 1 kbp. The trimmed metagenome reads were afterwards independently mapped to the assembled scaffolds using CLCs ‘map reads to reference’ algorithm, with a minimum similarity of 95% over 80% of the read length.

Open reading frames were predicted in the assembled scaffolds using the metagenome version of Prodigal14. A set of 107 HMMs of essential single-copy genes25 were searched against the predicted open reading frames using HMMER36 with default settings, except for the use of the trusted cut-off (-cut_tc). Identified proteins were taxonomically classified using BLAST against the RefSeq (version 52) protein database with a maximum e-value cut-off of 10⁻⁵. MEGAN27 was used to extract class level taxonomic assignments from the BLAST .xml output file. The script network.pl was used to extract paired-end read connections between scaffolds using a SAM file of the read mappings to the metagenome. Individual scaffolds were extracted using the multi-metagenome principles5 and refined using tetranucleotide frequencies, as implemented in the mmgenome R package32. The script extract.fastq.reassembly.pl was used to extract class A reads in preheated (48 °C) and diluted (1:1 with deionised water) film emulsion (Ilford K4). Absence of canonical bacterial or archaeal amoA sequences in the metagenome data was confirmed by searching a set of reference sequences against a BLAST database containing all trimmed metagenome reads.

Code availability. The Rmarkdown files used for extracting the genome bins are available for download52.

Activity assays. For activity assays, the reactor was supplied with medium containing labelled ammonium (1⁵⁷⁷ NH₄⁺). The medium flow was kept at normal operating rate (0.83 ml min⁻¹) and the biomass was stirred continuously. Isotopic composition of the ammonium and nitrite produced was measured by gas chromatography (Agilent 68900 equipped with a Porapak Q column at 80 °C and a TCD detector at 300 °C). Activity assays were conducted as described elsewhere47. For inhibition experiments ATU was added to a final concentration of 100 μM and biomass was reincubated for 10 min before substrate addition. Bottles were sealed with rubber stoppers and 10 ml air was added to the headspace to ensure slight overpressure. Incubations were performed at room temperature in the dark with mild agitation (50 rpm). At each time point, 0.5 ml sample was taken and stored at –20°C for further analysis.

Analytical methods. Ammonium was determined colorimetrically using a modification of the phosphotungstic-dialdehyde assay48 (detection limit 10μM) and nitrite (≥5μM) by the sulfanilamide reaction46. Nitrate (≥1μM) was determined using a method developed for NO₂⁻ and NO₃⁻ determination using a nitric oxide analyser (NOA280, GE Analytical Instruments, Manchester, UK). To determine the total organic carbon (TOC) concentration of the medium, first acidified to remove inorganic carbon. After 6.5 h dilution with ultrapure water, samples were measured using a TOC-L CPH/CPN analyser (Shimadzu, Düsseldorf, Germany).

Fluorescence in situ hybridization (FISH). For FISH analysis, samples from the reactor were fixed with 4% (v/v) paraformaldehyde (PFA), followed by hybridization with fluorescently labelled oligonucleotides as described elsewhere47. FISH probes used in this study (Extended Data Table 2) were 5’ labelled with the dyes FLUOS (5(6)-carboxyfluorescein-N-hydroxysuccinimide ester), Cy3 or Cy5 (Thermo Electron Corporation, Ulm, Germany). After hybridization, slides were air-dried and embedded in Vectashield (Vector Laboratories Inc., Burlingame, CA). Probe-conferred fluorescence was recorded on an Zeiss Axioplan 2 (Carl Zeiss AG, Oberkochen, Germany) equipped with a HBO 100 light source and specific filter sets for the detection of FLUOS, Cy3 and Cy5, or a Leica TCS SP2 AOBS (Leica Microsystems, Wetzlar, Germany) or a Zeiss LSM510 META (Carl Zeiss AG) confocal laser scanning microscope (CLSM), both equipped with one argon ion (450–514 nm) and two helium neon lasers (543 and 633 nm). Images were recorded with 63× glycerol or oil immersion objectives at a resolution of 1,024 × 1,024 pixels and 8-bit depth.

For quantifying relative biovolume fractions, PFA-fixed reactor biomass was hybridized to probes NspA662, Amx820 and EUB338mix (Extended Data Table 2) as described above. Subsequently, 45 image pairs were recorded at random fields of view using the Leica TCS SP2 AOBS CLSM. The images were imported into the image analysis software daim48 and evaluated as described elsewhere49.

AMO-labelling. Washed and disrupted (see above) biomass was incubated for 30 min with 15NH₄⁺ chloride and subsequently transferred to a fresh mineral medium and resuspended in 150 μl mineral medium containing no N-source. 12 ml biomass per incubation was transferred to 30 ml serum bottles and ammonium, nitrite or urea was added (200 μM final concentration). To test for ammonia activity and denitrification 1⁵⁷⁷ NH₄⁺ was used and the headspace analysed for labelled dinitrogen gas production as described above. For inhibition experiments ATU was added to a final concentration of 100 μM and biomass was reincubated for 10 min before substrate addition. Bottles were sealed with rubber stoppers and 10 ml air was added to the headspace to ensure slight overpressure. Incubations were performed at room temperature in the dark with mild agitation (50 rpm). At each time point, 0.5 ml sample was taken and stored at –20°C for further analysis.

FISH combined with microautoradiography (FISH-MAR). FISH-MAR experiments were performed as described before50. 150 ml biomass was taken from the reactor and filters were disrupted as described above. After harvesting and washing, the biomass was resuspended in mineral medium and transferred to serum bottles. Ammonium or nitrite was added to a final concentration of 500 μM. As controls, incubations with ammonium and ATU (100 μM), without nitrogen source and ammonium addition, were performed. PFA-fixed biomass was transferred to 150 μl CuCl₂ labelled bicarbonate were added to all samples, bottles were sealed with rubber stoppers and incubated at room temperature in the dark for 18 h. After incubation, the biomass was harvested by centrifugation (20,000g, 10 min), PFA-fixed and FISH was performed on coverslips as described above. Hybridized samples were dipped in preheated (48°C) and diluted (1:1 with deionised water) film emulsion (Ilford K4).
Nuclear Emulsion K5, Harman Technology, UK). After overnight drying at room temperature, samples were exposed for 6 days at 4°C and developed in Kodak D19 developer as described before. Images were recorded on a Zeiss LSM510 META CLSM as detailed above. To correct for the different levels of unspecific silver grain deposition in the incubations, the degree of silver grain formation in areas without biomass was compared to the amount of silver grains above biomass flocs. Only cell clusters which showed grain deposition clearly above background level were considered positive.

Phylogenetic analyses. 16S rRNA sequences with nucleotide identities $\geq$ 98% and amoA sequences with identities $\geq$ 70%, to the respective sequences of Nitrospira sp.1 or sp.2 were identified in the NCBI nr database by BLAST. 16S rRNA sequences were imported into the SILVA small subunit ribosomal RNA database release 119, amoA sequences in a custom-made database containing a reference set of amoA and pmoA sequences. nxrA sequences were imported in a custom-made database containing all published sequences from Nitrospira, Nitrospina and anammox organisms. Sequence alignments for all data sets where generated and manually refined using ARB 5.5. Bayesian inference trees were calculated using MrBayes 3.2.3 until a standard deviation $< 0.01$ was reached. For 16S rRNA analyses the GTR substitution model and a 50% conservation filter resulting in 1463 valid alignment positions were used. amoA genes were translated into their amino acid sequence and a 10% conservation filter resulting in 264 alignment positions in combination with the WAG substitution model were used for tree calculation. nxrA trees were calculated from nucleic acid sequences with the GTR substitution model and a 50% conservation filter resulting in 2,660 distinct alignment patterns. For trees were calculated from nucleic acid sequences with the GTR substitution model and a 50% conservation filter resulting in 1463 valid alignment positions were used. amoA genes were translated into their amino acid sequence and a 10% conservation filter resulting in 264 alignment positions in combination with the WAG substitution model were used for tree calculation. nxrA trees were calculated from nucleic acid sequences with the GTR substitution model and without conservation filter, resulting in 2,660 distinct alignment patterns. For all trees 50% majority rule consensus trees are shown.

Database mining. All 7,306 public shotgun metagenomes and metatranscriptomes available in MG-RAST were searched for the presence of the diagnostic amoA gene. Data sets were downloaded and searched against a small set of characteristic amoA sequences using DIAMOND with the default settings. The resulting 4,993 hits were filtered using a BLAST score ratio of the initial alignment score versus the NCBI nr.

51. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596 (2013).
52. Ludwig, W. et al. ARB: a software environment for sequence data. Nucleic Acids Res. 25, 263–269 (1996).
53. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference software. Bioinformatics 19, 1372–1374 (2003).
54. Rasko, D. A., Myers, G. S. & Ravel, J. Visualization of comparative genomic synteny results. Genome Res. 13, 1363–1371 (2003).
55. Wagner, M. & Stahl, D. A. Phylogenetic probes for analyzing abundance and spatial organization of anaerobic ammonia-oxidizing bacteria. Appl. Environ. Microbiol. 56, 1289–1297 (1999).
56. Rasko, D. A., Myers, G. S. & Ravel, J. Visualization of comparative genomic synteny results. Genome Res. 13, 1363–1371 (2003).
57. Schmid, M. et al. Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. Syst. Appl. Microbiol. 23, 93–106 (2000).
58. Schleifer, K.-H., Vennemann, T. D. & Deinema, a novel image analysis program for microbial ecology and biofilm research. Environ. Microbiol. 8, 200–213 (2006).
59. Schleifer, K.-H. & Wagner, M. Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. Appl. Microbiol. Biotechnol. 75, 237–248 (2007).
60. Lee, N. et al. Combination of fluorescent in situ hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. Appl. Environ. Microbiol. 65, 1289–1297 (1999).
61. Wagner, M., Rath, G., Amann, R., Koops, H.-P. & Schleifer, K.-H. 16S rRNA-targeted oligonucleotide probes for metagenomic surveying of microbial diversity. Appl. Environ. Microbiol. 62, 1186–1199 (2006).
62. Juretschko, S. et al. Combined molecular and conventional analyses of nitrosifying bacteria diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. Appl. Environ. Microbiol. 64, 3042–3051 (1998).
63. Wagner, M. et al. Design and application of fluorescent in situ oligonucleotide probes for direct quantification of anaerobic ammonia-oxidizing bacteria. Syst. Appl. Microbiol. 25, 263–269 (2002).
64. Wagner, M. et al. Design and application of fluorescent in situ oligonucleotide probes for direct quantification of anaerobic ammonia-oxidizing bacteria. Syst. Appl. Microbiol. 25, 263–269 (2002).
65. Wagner, M., Rath, G., Amann, R., Koops, H.-P. & Schleifer, K.-H. In situ identification of ammonia-oxidizing bacteria. Syst. Appl. Microbiol. 18, 251–264 (1995).
66. Juretschko, S. et al. Combined molecular and conventional analyses of nitrosifying bacterium diversity in activated sludge: Nitrosococcus mobilis and Nitrospira-like bacteria as dominant populations. Appl. Environ. Microbiol. 64, 3042–3051 (1998).
67. Moberly, B. K., Wagner, M., Urbain, V., Rittmann, B. E. & Stahl, D. A. Phylogenetic probes for analyzing abundance and spatial organization of nitrosifying bacteria. Appl. Environ. Microbiol. 62, 2156–2162 (1996).
68. Alm, E. W., Oerther, D. B., Larsen, N., Stahl, D. A. & Raskin, L. The oligonucleotide probe database. Appl. Environ. Microbiol. 62, 3557–3559 (1996).
Extended Data Figure 1 | Ammonium and nitrite conversion by the enrichment culture. a, b, Inorganic nitrogen load of the enrichment culture per 24 h cycle (filled symbols) and effluent concentrations (open symbols) for ammonium (a, diamonds) and nitrite (b, triangles). Effluent nitrite concentrations were below the detection limit (<5 μM) at all time points. Data points represent the mean of three technical replicates, error bars the standard deviations of these triplicates. Nitrate concentration in the medium varied between 0.5 and 2.0 mM and total organic carbon (TOC) content between 1.30 and 1.44 p.p.m., which was due to medium preparation with water obtained directly from the recirculation aquaculture system.
Extended Data Figure 2 | Metagenome binning. a, b. Extraction of the *Nitrospira* sp.1 (a) and sp.2 (b) genome sequences from the metagenome using differential coverage binning. Each circle represents a metagenomic scaffold, with size proportional to scaffold length; the plots contain a total of 47,584 scaffolds. The inlay of each figure shows the secondary binning based on tetranucleotide frequencies, with a total of 331 (a) and 281 (b) scaffolds included. Taxonomic classification is indicated by colour; a total of 3,158 essential marker genes were detected. The extracted bins are enclosed by a dashed line. c, d. Genome contaminations were excluded by generating linkage maps of the final bins of sp.1 (c, 25 scaffolds) and sp.2 (d, 86 scaffolds) using mate-pair sequencing data.
Extended Data Figure 3 | Phylogenetic analysis of NXR. Bayesian interference tree (s.d. = 0.0099) showing the affiliation of the *Nitrospira* sp.1 and sp.2 *nxrA* sequences in comparison to other genome-sequenced *Nitrospira*, *Nitrospina* and anammox bacteria. Posterior probabilities ≥70% and ≥90% are indicated by open and filled circles, respectively. NCBI protein accession numbers for all publicly available sequences are indicated, numbers with an asterisk are IMG gene IDs. The described *Nitrospira* sublineages are indicated by coloured boxes and roman numbers. The scale bar represents 10% sequence divergence. Note the different affiliation of the 'Candidatus *N. nitrosa* (sp.1) *nxrA* sequences. The tree contains 25 sequences from 12 species, belonging to 3 different phyla. Sequences from closely related bacterial putative nitrate reductases were used as outgroup (n = 4); the outgroup position is indicated by the arrow.
Extended Data Figure 4 | 16S rRNA-based phylogenetic analysis.
Bayesian interference tree (s.d. = 0.0098) showing the affiliation of the Nitrospira sp.1 and sp.2 16S rRNA sequences within Nitrospira sublineage II. Posterior probabilities ≥70% and ≥90% are indicated by open and filled circles, respectively. The strongly supported sequence group containing the novel Nitrospira spp. catalysing complete nitrification is shaded in grey, the two subgroups containing Nitrospira sp.1 and sp.2 (in bold) are highlighted by green and red boxes, respectively.

Extended Data Table 2 | Reactions of the newly designed FISH probe Ntspa476

© 2015 Macmillan Publishers Limited. All rights reserved
Extended Data Figure 5 | Control experiments of AMO-labelling.

a, Cells incubated with the fluorescent dye FTCP (green) were stained by FISH using probes specific for *Nitrospira* (Ntspa662, red) and all bacteria (EUB338mix, blue). A small cell cluster was stained by FTCP and targeted by both probes (resulting in a white overlay signal), while all other bacteria (in blue) were not or only slightly stained by FTCP. The green signal is due to autofluorescence and unspecific FTCP binding to the floc matrix.

b, Anammox cells (Amx820, blue) showed minor staining by FTCP (green), but to a much lesser degree than *Nitrospira* (Ntspa662, red; yellow overlay).

c and d, Positive controls: ammonium oxidizing bacteria (c, Nso1225 and Nso190, red) in an aerobic enrichment culture and a *Nitrosomonas europaea* pure culture (d, NEU, red, and EUB338mix, blue) were stained by FTCP (resulting in yellow and white overlays, respectively).

e and f, Negative controls: canonical *Nitrospira* in an aerobic enrichment culture (e, Ntspa662, blue) and a *Nitrospira moscoviensis* pure culture (f, Ntspa662, red, and EUB338mix, blue; magenta overlay) did not show any labelling with FTCP (green). The two bright green structures in (c) and the bright pink signal in (e) are due to autofluorescence. Images are representative of two (a and b) or one (c to f) individual experiments, with three technical replicates each. Scale bars in all panels represent 10 μm.
Extended Data Figure 6 | Batch incubations with nitrite, urea and without substrate. a, b, Nitrite (triangles) oxidation by the enrichment culture to nitrate (squares) in the absence (a) and in the presence (b) of ATU. The ammonia (diamonds) in b presumably stems from biomass decay and is not oxidized owing to ATU inhibition. c, Urea conversion to ammonium (diamonds) and subsequent oxidation to nitrate (squares). d, No-substrate control; minor amounts of ammonium (diamonds) presumably stem from mineralisation of degrading biomass, leading subsequently to nitrate (squares) formation. Symbols in all plots represent averages of three independent incubations; ammonium was determined in single measurements, nitrite and nitrate in duplicate (a and b) or triplicate (c and d). Error bars represent standard deviations of three biological replicates.
Extended Data Figure 7 | Ammonium and nitrite-dependent CO₂ fixation shown by FISH-MAR. 

a–d, FISH with probes for all bacteria (EUB338mix, blue), and probes specific for Nitrospira (Ntspa662, red; resulting in magenta) and anammox bacteria (Amx820, green; resulting in cyan). 

a. Ammonia-dependent carbon fixation. Only Nitrospira cells were active, as indicated by silver grain deposition. Note the inactive anammox cells on the left side of the smaller floc, co-localizing with highly active Nitrospira cells on the right side of the same floc. 

b. Inhibition of ammonia-dependent carbon fixation by ATU. 

c. Nitrite-dependent carbon fixation. Only Nitrospira cells incorporated ¹⁴CO₂. 

d. No-substrate control. Images are representative of two individual experiments, with two technical replicates each. Scale bars in all panels represent 10µm.
Extended Data Table 1 | General genomic characteristics of *Nitrospira* sp.1 and sp.2

| Bin                  | Ca. *N. nitrosa* (sp.1) | Ca. *N. nitrificans* (sp.2) |
|----------------------|-------------------------|-------------------------------|
|                      | initial | final | initial | final | initial | final |
| Genome size (bp)     | 4413075 | 4422398 | 4088547 | 4117083 |
| Contigs              | 25      | 15     | 86      | 36     |
| Largest contig (bp)  | 1073143 | 1804237 | 335390  | 475968 |
| N50                  | 659693  | 727365 | 103850  | 174194 |
| # Ns per 100 Kbp     | 355     | 0      | 420     | 0      |
| Completeness*        | 99% (97%) | >99% (97%) | >95% (97%) | >95% (97%) |
| Contamination*       | 0% (2.3%) | 0% (2.3%) | <1% (2.8%) | <1% (2.7%) |
| Coverage (CTAB)†     | 13.0    | 13.0   | 13.0    | 13.0   |
| Coverage (Kit)†      | 10.0    | ‡      | 5.0     | ‡      |
| Average G+C content  | 54.8    | ‡      | 56.6    | ‡      |
| Number of coding sequences (CDS) | 4309 | ‡ | 4502 | ‡ |
| rRNA operons         | 1       | ‡      | 1       | ‡      |
| tRNAs                | 46      | ‡      | 43      | ‡      |

*Values are based on evaluation of the binning plots and manual inspection; numbers in brackets are based on CheckM40.
†For details on DNA extraction see Methods section.
‡These values were only determined for the final genomic bins.
| Probe name | Probe full name* | Sequence (5'-3') | Binding position† | FA%‡ | Specificity | Ref. |
|------------|------------------|------------------|------------------|------|-------------|------|
| Amx820     | S-*-Amx-0820-a-A-22 | AAA ACC CCT CTA CTT AGT GCC C | 820 - 841 | 40 | Genera *Brocadia, Kuenenia* | 57  |
| Arch915    | S-D-Arch-0915-a-A-20 | GTG CTC CCC CGC CAA TTC CT | 915 - 934 | nd§ | Domain *Archaea* | 58  |
| Eub338I    | S-D-Bact-0338-a-A-18 | GCT GCC TCC CGT AGG AGT | 338 - 355 | 0 - 50 | Domain *Bacteria* | 59  |
| Eub338II†  | S-*-Bact-0338-b-A-18 | GCA GCC ACC CGT AGG TGT | 338 - 355 | 0 - 50 | Order *Phanerochaete* | 60  |
| Eub338III† | S-*-Bact-0338-c-A-18 | GCT GCC ACC CGT AGG TGT | 338 - 355 | 0 - 50 | Order *Verrucomicrobiales* | 60  |
| NEU        | S-*-Nsm-0651-a-A-18 | CCC CTC TGC TGC ACT CTA | 653 - 670 | 40 | *Nitrosomonas* spp. | 61  |
| cNEU       | -                 | TTC CAT CCC CCT CTG CCG | 659 - 676 | - | Competitor to NEU | 61  |
| NmV        | S-S-Nmob-0174-a-A-18 | TCC TCA GAG ACT ACG CCG | 174 - 191 | 35 | *Nitrosococcus mobilis* lineage¶ | 62  |
| Nso190     | S-F-bAOB-0189-a-A-19 | CGA TCC CCT GCT TTT CTC C | 189 - 207 | 55 | Betaproteobacterial AOB | 63  |
| Nso1225    | S-F-bAOB-1224-a-A-20 | CGC CAT TGT ATT ACG TGT GA | 1224 - 1243 | 35 | Betaproteobacterial AOB | 63  |
| Ntspa662   | S-G-Ntspa-662-a-A-18 | GGA ATT CCG CGG TCC TCT | 662 - 679 | 35 | Genus *Nitrospira* | 26  |
| cNtspa662  | -                 | GGA ATT CCG CTC TCC TCT | 662 - 679 | - | Competitor to Ntspa662 | 26  |
| Ntspa712   | S-*-Ntspa-712-a-A-21 | CGC CTT CGC CAC CGG CTC TCC | 712 - 732 | 35 | Phylum *Nitrospira* | 26  |
| cNtspa712  | -                 | CGC CTT CGC CAC CGG TCG TCC | 712 - 732 | - | Competitor to Ntspa712 | 26  |
| Ntspa476   | S-*-Ntspa-0476-a-A-22 | CTG CAG GTA CCG TCC GAA | 476 - 494 | 20 | Ca. *N. nitrosa, Ca. N. nitrificans* | This study |
| cNtspa476  | -                 | CTG GAG GTA CCG TCC GAA | 476 - 494 | - | Competitor to Ntspa476 | This study |

*Probe nomenclature according to Alm et al. See reference 64.
†Probe binding position according to *Escherichia coli* 16S rRNA gene numbering.
‡Percent formamide (v/v) added to the hybridization buffer for optimal hybridization stringency.
§Not determined.
¶Probes where used in a equimolar mixture (EUB338mix) to detect all *Bacteria*.
‖Probe targets *N. mobilis*, which is affiliated with the betaproteobacterial *Nitrospira* lineage and not the gammaproteobacterial genus *Nitrosococcus*.

References 57–64 are cited in this table.
## Extended Data Table 3 | Metagenome screening for *Nitrospira*-like amoA sequences

| Source                          | Geographical location          | Number of hits  | Total reads† | Project name                          | Dataset ID‡ |
|--------------------------------|--------------------------------|-----------------|--------------|---------------------------------------|-------------|
| **Metagenome projects**         |                                |                 |              |                                       |             |
| River sediment                  | Tongue river, Montana, USA     | 1327            | 556,961,375  | Tongue_all_2011                       | 4481956-57 ; 63-72 ; 74-86 |
| Soil                            | Houston, Texas, USA            | 367             | 321,988,632  | Metagenomic investigation for an ethanol-blended fuel spill | 4519753-58 ; 60-64 ; 67-76 |
| Prairie soil                    | Auburn, Illinois, USA          | 119             | 1,075,325,181| ISA-SMC-2011                          | 4502539-2541 ; 2543 ; 2923-2924 ; 2926 ; 2928 ; 2930 ; 2932-33 ; 2935 |
| Soil                            | Ha Noi, Vietnam                | 94              | 246,030,284  | Rice field                            | 4628743-47 ; 53-54 |
| Garden soil                     | Xiamen, Fujian, China          | 80              | 46,831,964   | Metagenome screening for unusual PmoA sequence group | 4635904-5 |
| Air                             | Beijing, China                 | 68              | 978,592,643  | Beijing PM2.5 and MP10 Pollutants      | 4516402-6403 ; 6455 ; 6459 ; 6637 ; 6651 ; 6802-6803 ; 6910-6911 ; 6952 ; 7064 |
| Agricultural soil               | Amazonia, Brazil               | 63              | 254,067,071  | Amazon Soil metagenome 2_mendes        | 4497370-371 ; 376 ; 391-393 ; 395-396 ; 407-409 ; 411-412 |
| Marine sediment                 | Gulf of Mexico, USA            | 45              | 2,425,926,864| BP_Sediments                          | 4510162-66 ; 68-69 ; 71 ; 73-74 |
| Marine sediment                 | Plum Island, Massachusetts, USA| 33              | 38,370,475   | IGERT Reverse Ecology 2011-2013        | 4519628 ; 19632 ; 19636 ; 20031 |
| Activated sludge                | Stanley wtp, Hong Kong         | 26              | 16,663,946   | Stanley wtp activated sludge sample    | 4467420     |
| Soil                            | Danum, Malaysia                | 24              | 43,344,688   | Effect of logging on soil microbial community in tropics | 4582264-267 ; 270 ; 798 ; 802-803 ; 805 |
| Agricultural soil               | Richmond, Indiana, USA         | 23              | 70,731,826   | EarlHamMetagenomes2012                | 4508937-38 ; 40 |
| wtp sludge                      | Malaysia                       | 23              | 40,000,000   | UTM waste water treatment project A    | 4544292-4293 ; 4301 ; 4307;5190; 6367-6368; 6370; 6373; 6375 |
| Activated sludge                | Switzerland                    | 20              | 9,455,087    | Swiss wtp metatranscriptomics          | 4491800     |
| Soil                            | Cologne, Germany               | 20              | 46,128,675   | Barley                                | 4529836 ; 30504 |
| Alkaline travertine water       | Voldr Massif, Liguria, Italy   | 19              | 42,594,481   | Microbial Biogeography of Serpentinites| 4537864-69 |
| Soil                            | Iowa, USA                      | 16              | 790,560,095  | GP corn unassembled                   | 4539519 ; 21 ; 23 ; 28 |
| Sports facility soil            | Norman, Oklahoma, USA          | 15              | 10,247,092   | Natural products                      | 4573678 ; 83 |
| River water                     | Minnesota, USA                 | 14              | 60,806,478   | M3P 2012                              | 4534334-35 ; 45-47 |
| Ochard soil                     | Haifa, Israel                  | 13              | 27,265,311   | Revital_qt                            | 4631721 ; 24 |
| Freshwater sediment             | Rifle, Colorado, USA           | 8               | 236,916,472  | Subsurface Rifle                      | 4465820 ; 4465822 |
| Rizosphere                      | Golm, Germany                  | 8               | 32,897,323   | Barley_Rhizomicrobiomics_test_B_PE    | 4524591-96 |
| Coral reef                      | Xisha island, China            | 8               | 125,160,089  | S_TS_MG                               | 4580696-698 ; 702 |
| Soil                            | Basque Country, Spain          | 6               | 3,293,845    | Metal_soil                            | 4510865     |
| Mine soil                       | Coto Txomin, Spain             | 6               | 196,440      | Pb-Zn-Mine                            | 4580863 ; 73 |
| River biofilm                   | West Virginia, USA             | 6               | 3,487,276    | MTR_GeMS_DNA                          | 4589540-1 |
| Marine sediment                 | Santa Barbara, California, USA | 5               | 96,123,985   | Scott_Nitro                           | 4537093     |
| Cave microbial mat              | Weebubble cave, Eucla, Australia| 4               | 475,608      | Weebubble Cave Slime Curtain Metagenome| 4448052    |
| Groundwater                     | Tulum, Quintana Roo, Mexico    | 4               | 59,482,508   | Yucatan Groundwater                   | 4536382-3   |
| Grassland soil                  | Bethel, Minnesota, USA         | 4               | 71,162,444   | CedarCreek_minsoil_june2013           | 4541645     |
| Soil                            | Amazonia, Brazil               | 3               | 23,648,292   | Amazon Soil metagenome 1              | 4493652     |
| Freshwater microbial mat        | Hot creek, Colorado, USA       | 3               | 6,877,377    | International geobiology course 2014 PreTrip | 4549766     |
| River sediment                  | Athabasca, Alberta, Canada     | 2               | 2,524,335    | Athabasca-biofilms                    | 4482887     |

### Metatranscriptome projects

| Source                          | Geographical location          | Number of reads | Project name                          | Dataset ID‡ |
|--------------------------------|--------------------------------|-----------------|---------------------------------------|-------------|
| River microbial mat             | West Virginia, USA             | 523             | 174,983,655 | MTR_GeMS_RNA                          | 4597881-86 |
| Oil contaminated soil           | Varennes, Quebec, Canada       | 164             | 234,156,703 | GenoRem_GH_MT                          | 4512573 ; 576-580 ; 586 ; 590 ; 592 ; 608 |
| Soil                            | Kalamazoo, Michigan, USA       | 28              | 205,252,966 | Miscanthus Metatranscriptome            | 4554103     |
| Marine sediment                 | Gulf of Mexico, USA            | 9               | 152,742,090 | MG-Core_Metat_Merged                   | 4508038 ; 41 ; 53 |
| Paddy soil                      | Jiangdu, China                 | 6               | 52,898,024  | paddy soil                             | 4553284-5   |

---

*Number of sequences affiliated with the novel AmoA/Unusual PmoA sequence group.
†Total number of metagenomic reads in the respective MG-RAST project.
‡For retrieving these datasets from MG-RAST '.3' must be added to the respective dataset ID.

© 2015 Macmillan Publishers Limited. All rights reserved