

Propagation of viral genomes by replicating ammonia-oxidising archaea during soil nitrification

Sungeun Lee¹, Ella T. Sieradzki¹, Graeme W. Nicol^{1*} and Christina Hazard^{1*}

¹ Univ Lyon, CNRS, INSA Lyon, Université Claude Bernard Lyon 1, Ecole Centrale de Lyon, Ampère, UMR5005, 69134 Ecully, France

* Corresponding authors. Email: graeme.nicol@ec-lyon.fr

christina.hazard@ec-lyon.fr

Abstract

Ammonia-oxidising archaea (AOA) are a ubiquitous component of microbial communities and dominate the first stage of nitrification in some soils. While we are beginning to understand soil virus dynamics, we have no knowledge of the composition or activity of those infecting nitrifiers or their potential to influence processes. This study aimed to characterise viruses having infected autotrophic AOA in two nitrifying soils of contrasting pH by following transfer of assimilated CO₂-derived ¹³C from host to virus via DNA stable-isotope probing and metagenomic analysis. Incorporation of ¹³C into low GC mol% AOA and virus genomes increased DNA buoyant density in CsCl gradients but resulted in co-migration with dominant non-enriched high GC mol% genomes, reducing sequencing depth and contig assembly. We therefore developed a hybrid approach where AOA and virus genomes were assembled from low buoyant density DNA with subsequent mapping of ¹³C isotopically enriched high buoyant density DNA reads to identify activity of AOA. Metagenome-assembled genomes were different between the two soils and represented a broad diversity of active populations. Sixty-four AOA-infecting viral operational taxonomic units (vOTUs) were identified with no clear relatedness to previously characterised prokaryote viruses. These vOTUs were also distinct between soils, with 42% enriched in ¹³C derived from hosts. The majority were predicted as capable of lysogeny and auxiliary metabolic genes included an AOA-specific multicopper oxidase suggesting infection may augment copper uptake essential for central metabolic functioning. These findings indicate virus infection of AOA may be a frequent process during nitrification with potential to influence host physiology and activity.

Main text

Microbially mediated oxidation of ammonia to nitrate during nitrification is a central component of the global nitrogen (N) cycle. It is also responsible for major losses of applied fertiliser N in soil, generating atmospheric pollution via direct and indirect production of nitrous oxide (N₂O) as well as nitrate (NO₃⁻) pollution of groundwater [1]. Autotrophic ammonia-oxidising archaea (AOA) of the class *Nitrososphaeria* are a ubiquitous component of soil microbial communities and often dominate ammonia oxidation and nitrification-associated N₂O emissions when ammonia is supplied at low rates via organic matter mineralisation [2], slow-release fertilisers [3] or in acidic soils [4]. Integrated temperate viruses (proviruses) and other virus-associated protein-encoding genes are found in most AOA genomes suggesting frequent interaction (see Supplementary Text). While viruses infecting marine AOA have been characterised through metagenomic approaches [5] and cultivation [6], those infecting soil AOA or other nitrifier groups are currently uncharacterised.

Virus infection can influence biogeochemical cycling by augmenting host activity or causing cell mortality and subsequent release of nutrients [7]. Recent advances have demonstrated that soil virus communities are dynamic in a wide range of soils [e.g. 8, 9] and augmenting virus loads modulate C and N fluxes [10, 11]. Nevertheless, identifying active interactions with specific populations or functional groups in soil remains challenging due to structural complexity and the vast diversity of hosts and viruses. Recent use of stable-isotope approaches has investigated whole community host-virus dynamics [12, 13] or interactions between individual host-virus populations specific to a functional process and substrate [14]. The aim of this study was to utilise the latter approach with ¹³CO₂-based DNA-SIP to focus on nitrification-associated interactions and to test the hypothesis that viruses are a dynamic component of soil AOA activity.

Hybrid analysis of GC mol% fractionation and ¹³C-DNA-SIP for identifying active AOA hosts and viruses

Microcosms were established (see Supplementary Text) with two sandy loam soils of contrasting pH (4.5 or 7.5), which were characterised previously as containing distinct AOA communities [15]. As ammonia from hydrolysed urea is used by AOA in these and other soils [1, 3], nitrification was fuelled by addition of urea (two pulses of 100 µg N g⁻¹ soil (dry weight)) to stimulate AOA growth and potential enrichment of AOA-infecting viruses. Microcosms were incubated at 25°C for 30 days (Supplementary Figure 1) with a 5% ¹²C- or ¹³C-CO₂ headspace to target viruses of autotrophs via transfer of assimilated carbon. Extracted DNA was subjected to isopycnic centrifugation in CsCl gradients, in which DNA migrates as a function of GC mol% and isotopic enrichment, before fractionation and analysis of genomic DNA of different buoyant densities [14]. Quantification of AOA genomic DNA distribution across CsCl gradients demonstrated growth of AOA populations (Figure 1A; Supplementary Figure 2) and genomic DNA in high buoyant density (HBD) fractions (>1.719 g ml⁻¹) was then pooled for each replicate microcosm and metagenomes sequenced (Supplementary Table 1). While ¹³C-incorporation clearly separated DNA from growing vs non-growing AOA populations, ¹³C-enriched AOA genomes (average contig GC mol% content of 38%) co-migrated within the CsCl gradient with more abundant community ¹²C genomes with higher GC mol% (>57%), reducing AOA host and virus sequencing depth, contig assembly and binning. Consequently, only one medium-quality AOA metagenome-assembled genome (MAG) was recovered from HBD DNA and no AOA-associated virus contigs were identified amongst 290 viral operational taxonomic units (vOTUs) ≥10 kb (Figure 1B).

To facilitate analysis of low GC mol% genomes while also identifying activity, we developed a hybrid approach combining metagenomic analysis of low buoyant density (LBD) ^{12}C DNA with read-mapping of sequenced HBD DNA (Figure 1; Supplementary Figure 2). The AOA and virus genomes were assembled from metagenomic libraries of ^{12}C LBD DNA only ($<1.699 \text{ g ml}^{-1}$) representing both dormant and active populations. Individual sequence reads from HBD DNA libraries from both ^{12}C - and ^{13}C -incubations were then mapped to identify activity with a significantly greater relative abundance in ^{13}C -incubated samples demonstrating isotope incorporation. For taxonomically defined contigs of hosts or host-linked viruses, in comparison to HBD, assemblies from LBD libraries increased the relative proportion of contigs from the AOA-containing *Thermoproteota* phylum (GTDB taxonomy; *Nitrososphaerota* in ICNP taxonomy and *Thaumarchaeota* in NCBI taxonomy) and *Thermoproteota*-infecting virus contigs from 0.3 to 2.8% and 0 to 22.3%, respectively (Supplementary Figure 3). In LBD metagenomes, 123 medium- and high-quality MAGs from both soils were assigned taxonomy (Supplementary Table 2) including nine AOA MAGs representing six genera (GTDB classification [16]) within the *Nitrososphaerales* and *Nitrosopumilales* orders and one nitrite-oxidising bacterium (NOB) *Nitrobacter* MAG (MAG 15) (Figure 1B, Supplementary Figure 4), contributing to the oxidation of nitrite (NO_2^-) to NO_3^- during the second step of canonical nitrification. Mapping of HBD sequence reads revealed that eight of the ten MAGs were derived from active autotrophic populations (Figure 1C). Consistent with previous studies characterising the distribution of soil AOA from different taxa [17], this included four *Nitrosotalea* genus populations (MAGs 9, 61, 77 and 107) from pH 4.5 soil and three *Nitrososphaera* genus populations (MAGs 12, 52 and 96) from pH 7.5 soil. No evidence of autotrophic growth was observed for populations represented by a further two MAGs within the *Nitrososphaerales*. Both are from genera with no cultivated

representatives with MAG 112 from genus TA-21 (*amoA*-gene lineage NS- β -2 [17]) and MAG 63 (*amoA*-gene lineage NS- δ -2.1) representing a novel genus (designated here as CS63). The latter belongs to the *amoA* gene-defined NS- δ lineage that while widely found in soil, is rarely implicated in contributing to ammonia oxidation. Although these two MAGs may simply be from AOA populations not active under the incubation conditions used, they are also consistent with microorganisms possessing physiologies distinct from cultivated AOA.

AOA viruses

After predicting contigs of virus origin using established bioinformatic tools, those from AOA-infecting viruses were identified with a custom database of hallmark genes from AOA proviruses and analysis of other shared homologues (see Supplementary Text). Specifically, proviruses were identified in 184 (87%) of 212 AOA genome sequences downloaded from GTDB. Capsid, terminase, portal and integrase protein sequences from these proviruses were subsequently used as search queries. While previous analysis of 12 total soil metagenome or virome libraries from the same soil samples did not enable identification of any AOA-associated virus contigs (average of 152 million quality filtered reads and 1,947 contigs ≥ 10 kb per library) [14], this targeted approach using LBD DNA from nitrifying microcosms yielded 64 contigs predicted as derived from AOA-infecting viruses and each representing a unique vOTU. In addition, viruses infecting other nitrifier populations were identified with five and three vOTUs derived from viruses infecting ammonia-oxidising bacteria (AOB) and NOB, respectively. Virus and host genome GC mol% correlate [18] and AOA-infecting virus contigs possessed a similar average content of 36.3% which was considerably lower than averages of 46.6% and 67.0% of all viral contigs recovered from LBD and HBD

metagenomes, respectively. As with their hosts, AOA virus communities were distinct
 140 between the two pH soils (Figure 2A) as also observed previously for total virus
 communities [19] with 29 (45%) and 34 (53%) vOTUs significantly more relatively
 142 abundant in pH 4.5 and 7.5 LBD libraries compared to the other soil, respectively.
 Comparison of the relative abundance of HDB sequence reads from both ^{12}C and ^{13}C
 144 libraries revealed that 27 contigs from both soils (42%) were enriched in ^{13}C and thus
 confirming infection during incubation and transfer of C from AOA hosts to viruses.

146 Soil AOA viruses were distinct from characterised archaeal and bacterial
 viruses. Gene-sharing network analysis using vConTACT v2.0 [20] did not identify any
 148 associations with reference sequences or recently described viruses of marine AOA
 [6] and genome-wide sequence comparisons [21] placed 60 of the 64 soil AOA viral
 150 contigs within a unique proteomic grouping (Figure 2B). The majority were predicted
 as capable of lysogeny with 55 encoding an integrase. While it can be difficult to
 152 determine virulent versus temperate capability from partial genomes, only 6.8 and
 8.9% of all virus contigs from LBD and HBD metagenomes possessed a predicted
 154 integrase. Although suggesting a higher proportion of AOA viruses may be temperate,
 it should be noted that this could be a consequence of using AOA provirus-associated
 156 genes to identify viral contigs, biasing towards identification of those from temperate
 viruses. These results also highlight that while DNA-SIP of total metagenomes
 158 enables identification of ^{13}C incorporation into virus genomes, it is potentially limited
 in facilitating confident bioinformatic prediction of free temperate virus particles versus
 160 proviruses present from a previous infection of a currently active host. While the
 integration status of a temperate virus can be predicted bioinformatically, using
 162 targeted “viromes” metagenomes of filtered virus particles will provide greater
 confidence of identifying genomes from free virus particles.

Eleven contigs contained large terminase (TerL) sub-unit gene sequences and were similar to those recovered from other predicted AOA virus or provirus genomes only (e-value $<10^{-5}$) (Figure 2C). In addition to viral hallmark genes encoding portal, endonuclease, tail and capsid proteins, contig 053596 contained a type 1 multicopper oxidase (MCO1) homologue found exclusively in AOA genomes [22] (Figure 2D). Copper (Cu) is essential for ammonia oxidation and electron transport in AOA and MCO1 is upregulated under Cu limitation and potentially involved in Cu sequestration from the environment [23]. This virus-encoded MCO1 homologue therefore represents an auxiliary metabolic gene that may contribute to a core process in AOA physiology.

Summary

These results from two soils indicate that virus infection of AOA during nitrification is likely a common and dynamic process, with the potential to influence nitrification rates in soil. Although viruses were only identified from contigs representing partial viral genomes, analysis of individual marker genes and genome-wide similarity comparisons revealed that those infecting AOA are highly divergent from known viruses. While one *Nitrobacter* MAG was obtained and both AOB and NOB viruses were also predicted, the hybrid approach developed in this study targeted low GC mol% DNA and was specifically tailored for the analysis of AOA populations and their viruses. A broader analysis of all viruses infecting nitrifying prokaryotic populations would therefore likely require a more comprehensive sequencing effort of all fractions throughout the CsCl gradient using quantitative SIP metagenomics [24] or alternative approaches, such as virus-targeted metagenomic viromes.

Data availability

Metagenome sequence data are deposited in NCBI's GenBank under BioProject

accession nos. PRJNA621418-PRJNA621429 (¹³C data) and PRJNA868779 (¹²C

data) with MAGs available under accession numbers SAMN30436125-

SAMN30436246.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was funded by an AXA Research Chair in Microbial Ecology awarded to GWN, Agence Nationale de la Recherche grant 'CONSERVE' (ANR-22-CE02-0006-

01) awarded to GWN and CH, and an EU Horizon Europe MSCA

Postdoctoral Fellowship 'DIVOBIS' awarded to ETS. Sequencing data generated

under JGI Community Science Program Proposal No. 503702 awarded to GWN and

CH was conducted by the US DOE JGI, a DOE Office of Science User Facility, and

supported by the Office of Science of the US DOE under Contract No. DE-AC02-

05CH11231. The authors would like to thank Dr. Willm Martens-Habbena (University

of Florida) for continued discussions on soil AOA ecophysiology and taxonomy, Dr.

Melina Kerou (University of Vienna) for guidance on AOA MCO diversity, Dr. Gary

Trubl for comments on the manuscript and Dr. Robin Walker (SRUC, Aberdeen) for

access to the much-lamented Craibstone Estate Woodlands Field long term

experiment.

Author contributions

SL, GWN and CH conceived and designed the experiment with additional input from

ETS. SL performed all laboratory work and analysed the data with input from GWN in

phylogenetic analyses. SL, GWN and CH drafted the manuscript text and figures with
all authors contributing to the final paper.

References

1. Prosser JL, Hink L, Gubry-Rangin C, Nicol GW. Nitrous oxide production by ammonia oxidizers: Physiological diversity, niche differentiation and potential mitigation strategies. *Glob Chang Biol.* 2020;26:103–118.
2. Huang L, Chakrabarti S, Cooper J, Perez A, John SM, Daroub SH, et al. Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. *ISME Commun.* 2021;1:19.
3. Hink L, Gubry-Rangin C, Nicol GW, Prosser JL. The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. *ISME J.* 2018;12:1084–1093.
4. Li Y, Chapman SJ, Nicol GW, Yao H. Nitrification and nitrifiers in acidic soils. *Soil Biol Biochem.* 2018;116:290–301.
5. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode *amoC* nitrification genes. *ISME J.* 2019;13:618–631.
6. Kim J-G, Kim S-J, Cvirkaite-Krupovic V, Yu W-J, Gwak J-H, López-Pérez M, et al. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc Natl Acad Sci U S A.* 2019;116:15645–15650.
7. Emerson JB. Soil Viruses: A New Hope. *mSystems.* 2019;4:e00120-19.
8. Santos-Medellín C, Estera-Molina K, Yuan M, Pett-Ridge J, Firestone MK, Emerson JB. Spatial turnover of soil viral populations and genotypes overlain by cohesive responses to moisture in grasslands. *Proc Natl Acad Sci U S A.* 2022;119:e2209132119.

9. Wu R, Davison MR, Gao Y, Nicora CD, Mcdermott JE, Burnum-Johnson KE, et
242 al. Moisture modulates soil reservoirs of active DNA and RNA viruses. *Commun
Biol.* 2021;4:992.
- 244 10. Braga LPP, Spor A, Kot W, Breuil M-C, Hansen LH, Setubal JC, et al. Impact of
phages on soil bacterial communities and nitrogen availability under different
246 assembly scenarios. *Microbiome.* 2020;8:52.
11. Albright MBN, Gallegos-Graves LV, Feeser KL, Montoya K, Emerson JB, Shakya
248 M, et al. Experimental evidence for the impact of soil viruses on carbon cycling
during surface plant litter decomposition. *ISME Commun.* 2022;2:1–8.
- 250 12. Starr EP, Shi S, Blazewicz SJ, Koch BJ, Probst AJ, Hungate BA, et al. Stable-
isotope-informed, genome-resolved metagenomics uncovers potential cross-
252 kingdom interactions in rhizosphere soil. *mSphere.* 2021;6:e0008521.
13. Trubl G, Kimbrel JA, Lique-Gonzalez J, Nuccio EE, Weber PK, Pett-Ridge J, et
254 al. Active virus-host interactions at sub-freezing temperatures in Arctic peat soil.
Microbiome. 2021;9:208.
- 256 14. Lee S, Sieradzki ET, Nicolas AM, Walker RL, Firestone MK, Hazard C, et al.
Methane-derived carbon flows into host–virus networks at different trophic levels
258 in soil. *Proc Natl Acad Sci U S A.* 2021;118:e2105124118.
15. Nicol GW, Leininger S, Schleper C, Prosser JL. The influence of soil pH on the
260 diversity, abundance and transcriptional activity of ammonia oxidizing archaea
and bacteria. *Environ Microbiol.* 2008;10:2966–2978.
- 262 16. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–
264 1927.

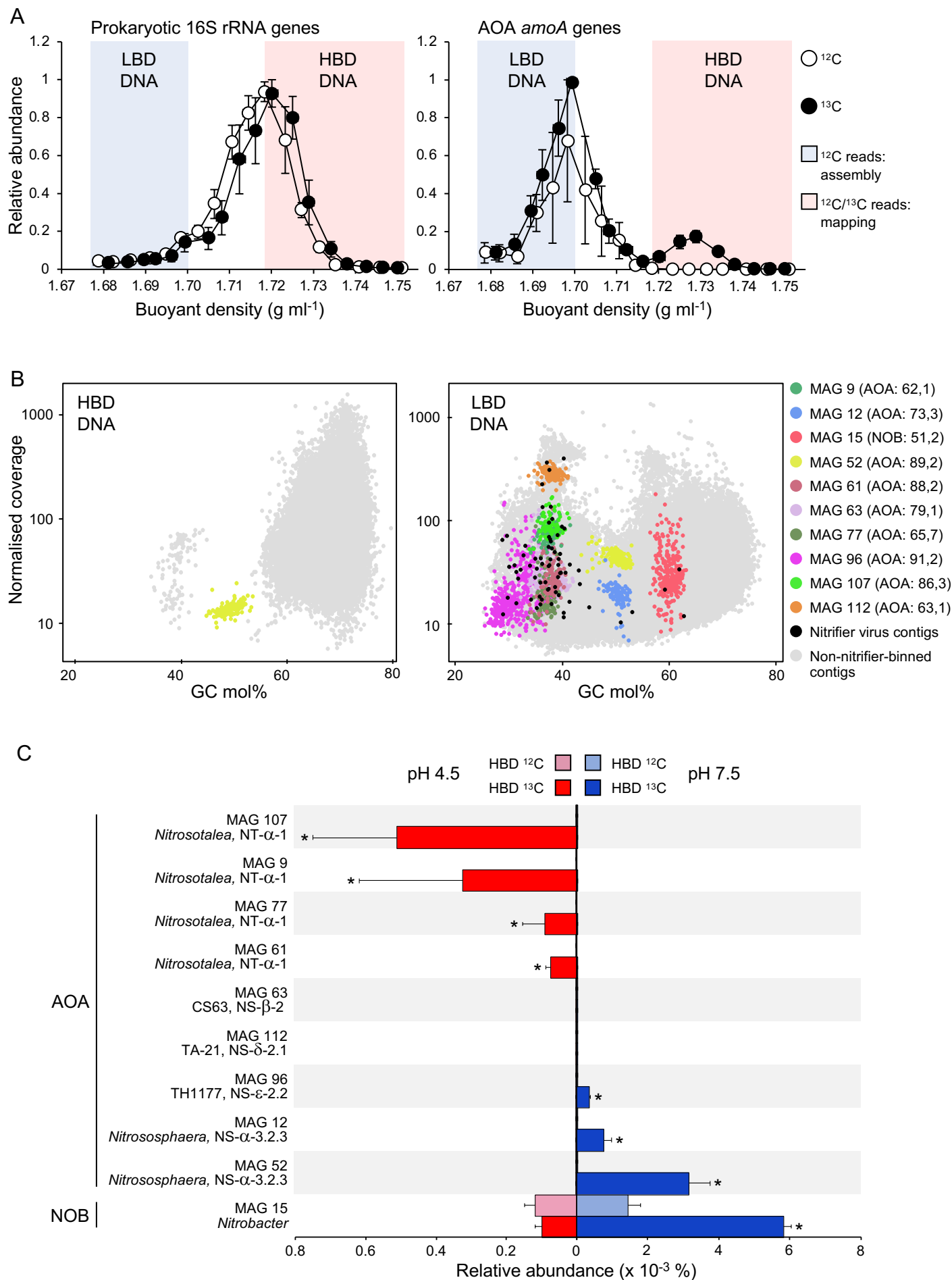
17. Alves RJE, Eloy Alves RJ, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying
the global phylogeny and environmental distribution of ammonia-oxidising
archaea based on amoA genes. *Nature Commun.* 2018;9:1517.
18. Cardinale DJ, Duffy S. Single-stranded genomic architecture constrains optimal
codon usage. *Bacteriophage.* 2011;1:219–224.
19. Lee S, Sorensen JW, Walker RL, Emerson JB, Nicol GW, Hazard C. Soil pH
influences the structure of virus communities at local and global scales. *Soil Biol
Biochem.* 2022;166:108569.
20. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al.
Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by
gene-sharing networks. *Nat Biotechnol.* 2019;37:632–639.
21. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the
viral proteomic tree server. *Bioinformatics.* 2017;33:2379–2380.
22. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, et al. Proteomics
and comparative genomics of *Nitrososphaera viennensis* reveal the core genome
and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci U S A.*
2016;113:7937–7946.
23. Reyes C, Hodgskiss LH, Kerou M, Pribasnig T, Abby SS, Bayer B, et al. Genome
wide transcriptomic analysis of the soil ammonia oxidizing archaeon
Nitrososphaera viennensis upon exposure to copper limitation. *ISME J.*
2020;14:2659–2674.
24. Sieradzki ET, Greenlon A, Nicolas AM, Firestone MK, Pett-Ridge J, Blazewicz SJ,
et al. Functional succession of actively growing soil microorganisms during
rewetting is shaped by precipitation history. *bioRxiv.* 2022;2022.06.28.498032.

Figure legends

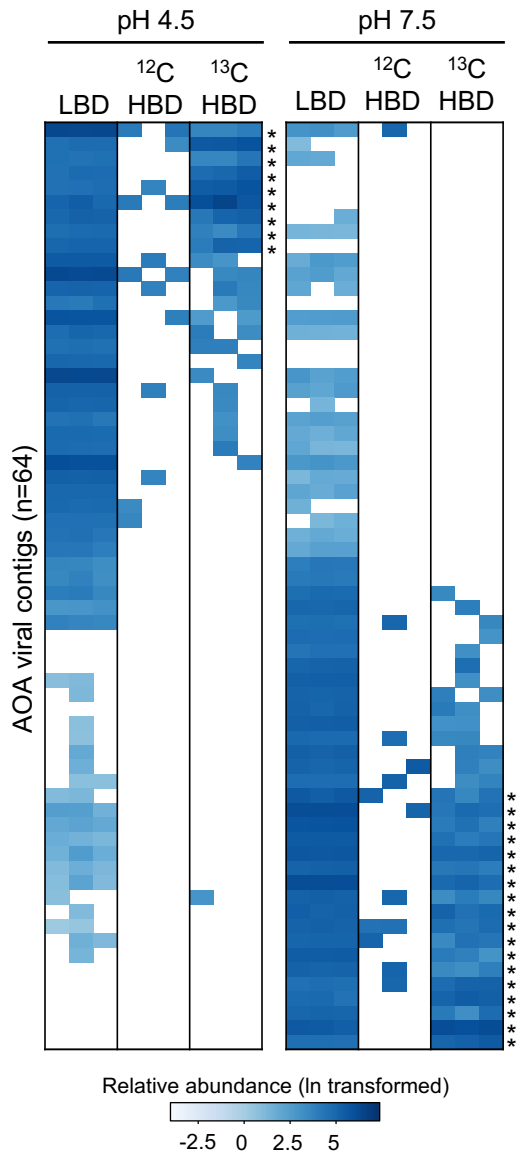
Figure 1. Contigs and MAGs from low GC mol% DNA and *in situ* growth of AOA in pH 4.5 and 7.5 soils determined from hybrid analysis of ^{12}C - and ^{13}C -enriched DNA. **A** Distribution of total prokaryotic and AOA genomes in CsCl gradients determined from the relative abundance of 16S rRNA genes and *amoA* genes, respectively. Example profiles are shown for pH 4.5 soil only (see Supplementary Figure 2 for pH 7.5 profiles). Vertical error bars are the standard error of the mean relative abundance and horizontal bars (mostly smaller than the symbol size) the standard error of the mean buoyant density of individual fractions from three independent CsCl gradients, each representing an individual microcosm. Genomic DNA in fractions highlighted in blue or pink areas were pooled for each replicate microcosm for metagenomic sequencing. **B** Distribution of assembled contigs (≥ 5 kb) from metagenomic libraries prepared from LBD or HBD DNA in sequence coverage vs. GC mol% plots. Contigs binned into ten medium- and high-quality nitrifier MAGs and predicted viruses of nitrifiers (≥ 10 kb) are highlighted, with number in parenthesis beside each MAG identifier giving estimated completeness and contamination (%). **C** Mean and standard error of the relative proportion of metagenomic reads from ^{12}C - and ^{13}C -enriched HBD DNA mapped onto contigs of nine AOA and one *Nitrobacter* MAGs. A significantly greater relative proportion of reads in ^{13}C -derived libraries are indicated with * ($p < 0.05$, two-sample Student's t-test or Welsch's t-test when variances were not homogenous). GTDB genus [16] and *amoA* gene lineage-affiliation [17] are given.

Figure 2. Identification and determination of activity of viruses infecting AOA in pH 4.5 and 7.5 soils using a hybrid analysis of ^{12}C - and ^{13}C -enriched DNA. **A** Heatmap showing the relative abundance of 64 vOTUs identified in LBD DNA libraries from pH 4.5 and 7.5 soil ($\geq 1\times$ coverage, $\geq 75\%$ contig breadth). To identify activity, reads from

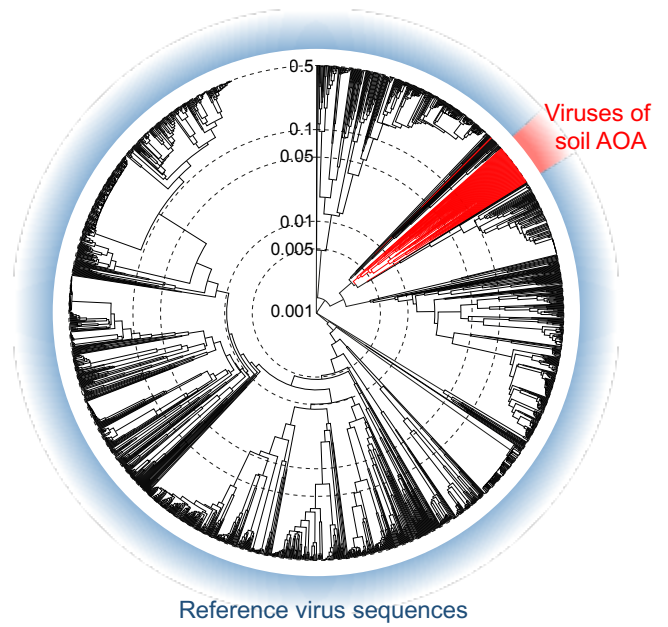
316 ^{12}C - and ^{13}C -enriched HBD DNA were mapped onto LBD-derived contigs with a
significantly greater relative proportion of reads in ^{13}C -derived libraries indicated with
318 * ($p < 0.05$, two-sample Student's t-test or Welsch's t-test when variances were not
homogenous). **B** Proteomic tree showing genome-wide sequence similarities between
320 AOA viral contigs and 1,981 curated virus genomes. Values at dotted circles represent
a distance metric based on normalized tBLASTx scores plotted on a log scale. **C**
322 Maximum likelihood phylogenetic analysis of derived large sub-unit terminase (TerL)
protein sequences (336 unambiguously aligned positions, LG substitution model)
324 identified in this and other metagenomic studies or in proviruses of cultivated AOA.
NCBI accession numbers are given in parenthesis. Circles at nodes represent
326 percentage bootstrap support from 1000 replicates, scale bar denotes an estimated
0.05 changes per position and colour-coding describes the comparative relative
328 abundance in each soil. **D** Genetic map of AOA-infecting virus-derived contig 053596
containing a type 1 multicopper oxidase (MCO1) gene. Maximum likelihood
330 phylogenetic tree describes placement of MCO1 within the AOA MCO/NirK family of
cultivated AOA following designations of Kerou *et al.* [22] (206 unambiguously aligned
332 positions, LG substitution model, invariant and gamma distributed sites). NCBI
accession numbers are given in parenthesis. Circles at nodes represent percentage
334 bootstrap support from 1000 replicates and the scale bar denotes an estimated 0.05
changes per position.



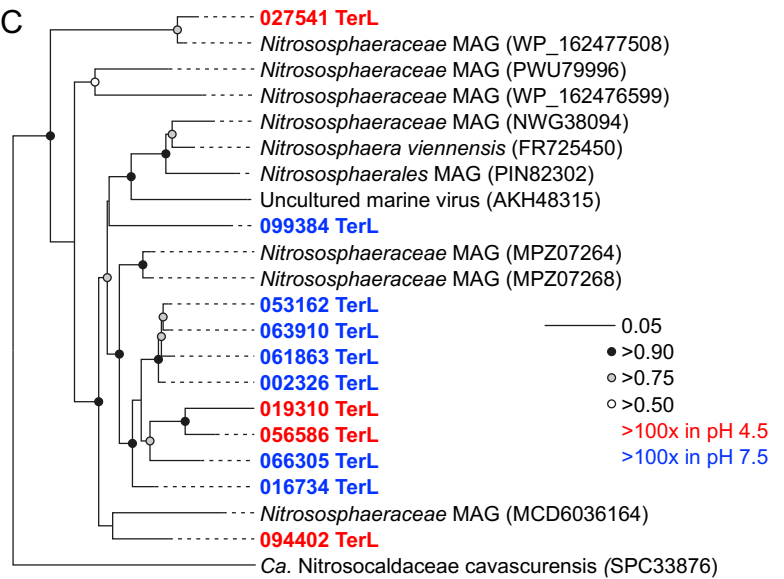
A



B



C



D

