



HAL
open science

Viruses of soil ammonia oxidising archaea identified using a novel DNA stable isotope probing approach for low GC mol% genomes

Sungeun Lee, Ella T Sieradzki, Christina Hazard, Graeme W Nicol

► **To cite this version:**

Sungeun Lee, Ella T Sieradzki, Christina Hazard, Graeme W Nicol. Viruses of soil ammonia oxidising archaea identified using a novel DNA stable isotope probing approach for low GC mol% genomes. 2022. hal-03758945

HAL Id: hal-03758945

<https://hal.science/hal-03758945v1>

Preprint submitted on 23 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2 **Viruses of soil ammonia oxidising archaea identified using a novel DNA stable**
isotope probing approach for low GC mol% genomes

4

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

6

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

10 * Corresponding authors. Email: christina.hazard@ec-lyon.fr;
graeme.nicol@ec-lyon.fr

12

14 **Abstract**

16 Ammonia oxidising archaea (AOA) are a ubiquitous component of microbial
18 communities which can dominate ammonia oxidation in some soils. While we are
beginning to understand soil virus dynamics, we have no knowledge of the
composition or activity of those infecting AOA or their potential to influence processes.
20 This study aimed to identify viruses infecting autotrophic AOA in two soils (pH 4.5 and
7.5) by following transfer of assimilated CO₂-derived ¹³C via DNA stable isotope
probing in CsCl isopycnic buoyant density (BD) gradients and metagenomic analysis.
22 Incorporation of ¹³C into low GC mol% AOA genomes (~35-40%) increased DNA BD
and separation from unenriched DNA, but also resulted in co-migration with dominant
24 non-¹³C-enriched high GC mol% bacterial genomes (>57%) reducing AOA
sequencing depth and contig assembly. We therefore developed a hybrid approach
26 where AOA and virus genomes were assembled from metagenomes from low BD
CsCl fractions from ¹²C-CO₂ incubations with subsequent mapping of high BD ¹³C- vs
28 ¹²C-enriched non-assembled reads to identify activity. AOA metagenome assembled
genomes (MAGs) were distinct between the two soils and represented a broad
30 diversity of active lineages. Sixty-four viral OTUs infecting AOA were also distinct
between soils, with 42% enriched in ¹³C. Comparative analyses demonstrated that soil
32 AOA viruses were distinct from characterised bacterial or archaeal viruses and
auxiliary metagenomic viral genes included an AOA-specific multicopper oxidase
34 involved in copper uptake essential for AOA metabolism. These observations in soils
with distinct AOA communities indicate that virus infection of AOA is likely a frequent
36 process during soil nitrification.

38

Main text

40 Autotrophic ammonia-oxidising archaea (AOA) of the class *Nitrososphaeria* are a
ubiquitous component of soil microbial communities and often dominate the first step
42 of nitrification and nitrification-associated N₂O emissions when ammonia is supplied
at low rates via organic matter mineralisation [1], slow-release fertilisers [2] or in acidic
44 soils [3]. Prophages and other virus-associated protein-encoding genes are abundant
in AOA genomes suggesting frequent interaction. While viruses infecting marine AOA
46 have been characterised through metagenomic approaches [4] and cultivation [5],
those infecting soil AOA or other nitrifier groups are currently uncharacterised.

48 Recent advances have demonstrated that soil virus communities are dynamic
in a wide range of soils [e.g. 6, 7] and augmenting virus loads modulate C and N fluxes
50 [8, 9]. Nevertheless, identifying active interactions with specific populations or
functional groups remains challenging due to the vast diversity of both host and viruses
52 in soil. Recent use of stable-isotope approaches has investigated whole community
host-virus dynamics [10, 11] or interactions between individual host-virus populations
54 specific to a functional process and substrate [12]. The aim of this study was to utilise
the latter approach with ¹³CO₂-based DNA-SIP to focus on nitrification-associated
56 interactions and to test the hypothesis that viruses are a dynamic component of soil
AOA activity.

58

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

Microcosms were established (see Supplementary Text) with two sandy loam soils of
62 contrasting pH (4.5 or 7.5) which were characterised previously as containing distinct
AOA communities [13]. Nitrification was stimulated by addition of urea (two pulses of

64 100 $\mu\text{g N g}^{-1}$ soil (dry weight)) and incubated at 25°C for 30 days (Supplementary
Figure 1) with a 5% ^{12}C - or ^{13}C -CO₂ headspace to target viruses of autotrophic nitrifiers
66 via transfer of assimilated carbon. DNA was extracted and subject to isopycnic
centrifugation in CsCl gradients, in which DNA migrates as a function of GC mol% and
68 isotopic enrichment, before fractionation and analysis of genomic DNA of different
buoyant densities [12]. Quantification of AOA genomic DNA distribution across CsCl
70 gradients demonstrated growth of AOA populations (Figure 1A; Supplementary Figure
2B) and genomic DNA in high buoyant density (HBD) fractions ($>1.719 \text{ g ml}^{-1}$) was
72 then pooled for each replicate microcosm and metagenomes sequenced
(Supplementary Table 1). While ^{13}C -incorporation clearly separated DNA from
74 growing vs non-growing AOA populations, ^{13}C -enriched genomes (average AOA
contig GC mol% content 38%) co-migrated within the CsCl gradient with community
76 ^{12}C genomes with higher GC mol% ($>57\%$), reducing AOA host and virus sequencing
depth, contig assembly and binning. Consequently, only one medium quality AOA
78 MAG and no AOA-associated virus contigs $\geq 10 \text{ kb}$ were identified (Figure 1B).

To facilitate analysis of low GC mol% genomes while also identifying activity,
80 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;
82 Supplementary Figure 2A). AOA and virus genomes were assembled from
metagenomic libraries of ^{12}C LBD DNA only ($<1.699 \text{ g ml}^{-1}$) representing both dormant
84 and active populations. Individual sequence reads from HBD DNA libraries from both
 ^{12}C - and ^{13}C -incubations were then mapped to identify activity. In comparison to HBD,
86 assemblies from LBD libraries increased the relative proportion of contigs from the
AOA-containing *Thaumarchaeota* phylum (NCBI taxonomy (*Thermoproteota* GTDB
88 taxonomy)) and *Thaumarchaeota*-infecting virus contigs from 0.3 to 2.8% and 0 to

22.3%, respectively (Supplementary Figure 3). In LBD metagenomes, 123 medium-
90 and high-quality MAGs from both soils were identified (Supplementary Table 2)
including 9 AOA MAGs representing 6 genera (GTDB classification [14]) within the
92 *Nitrososphaerales* and *Nitrosopumilales* orders and one nitrite oxidising bacterium
(NOB) *Nitrobacter* MAG (MAG 15) (Figure 1B, Supplementary Figure 4). Mapping of
94 HBD sequence reads revealed that 8 of the 10 MAGs were derived from active
autotrophic populations (Figure 1C). Consistent with previous studies characterising
96 the distribution of soil AOA from different taxa [15], this included 3 *Nitrosotalea* genus
MAGs from pH 4.5 soil and 4 *Nitrososphaera* genus MAGs from pH 7.5 soil. No
98 evidence of autotrophic growth was observed for a further two AOA MAGs within the
Nitrososphaerales. Both represented populations with no cultivated representatives at
100 the genus level with MAG 112 representing TA-21 (*amoA*-gene lineage NS- β -2) [15]
and MAG 63 representing a novel genus (designated here as CS63) (*amoA*-gene
102 lineage NS- δ -2.1). The latter is a representative of the *amoA* gene-defined NS- δ -
lineage which, while widely found in soil, is rarely implicated in contributing to ammonia
104 oxidation. While the recovery of these two MAGs may simply be representative of AOA
populations not active under the incubation conditions used, they are also consistent
106 with organisms possessing physiologies distinct from cultivated AOA.

108 *Active AOA viruses*

After predicting contigs of virus origin using established bioinformatic tools, those from
110 AOA-infecting viruses were identified with a custom database of hallmark genes from
AOA prophages and analysis of other shared homologues (see Supplementary Text).
112 Sixty-four contigs (each representing a different viral operational taxonomic unit
(vOTU)) were predicted as derived from AOA-infecting viruses in LBD DNA

114 metagenomic libraries. In addition, 5 and 3 contigs were derived from viruses infecting
ammonia oxidising bacteria (AOB) and NOB, respectively. Virus and host genomes
116 typically have similar GC mol% [16] and AOA-infecting virus contigs (average 36%)
were similar to their hosts (38%). As with their hosts, AOA virus communities were
118 distinct between the two pH soils (Figure 2A) as also observed previously for total virus
communities [17] with 29 and 34 vOTUs significantly more relatively abundant in pH
120 4.5 and 7.5 LBD libraries compared to the other soil, respectively. Mapping of HDB
sequence reads revealed that 42% of contigs were enriched in ^{13}C and thus confirming
122 infection during incubation and transfer of C from AOA hosts to viruses.

Analysis of genome-wide sequence comparisons using translated open reading
124 placed 60 of the 64 soil AOA viral contigs within one cluster but displayed no similarity
to recently described viruses of marine AOA [5] or characterised non-AOA archaeal
126 and bacterial viruses (Figure 2B). Eleven translated large terminase (TerL) sub-unit
gene sequences found in contigs were similar to those recovered from other predicted
128 AOA virus or prophage genomes only (e-value $<10^{-5}$) (Figure 2C). In addition to
containing viral hallmark genes encoding portal, endonuclease, tail and capsid
130 proteins, contig VC_053596 contained a type 1 multicopper oxidase (MCO1)
homologue [18] (Figure 2D) likely involved in the sequestration of copper from the
132 environment which is essential for ammonia oxidation and electron transport in AOA
[19]. As MCO1 is found exclusively in AOA genomes [18] it may represent a viral
134 auxiliary metabolic gene contributing to a core process in AOA physiology.

136 *Summary*

These results from two soils with distinct AOA communities indicate that virus infection
138 of AOA during soil nitrification is likely a common and dynamic process. Although

viruses were only identified from contigs representing partial viral genomes, analysis
140 of individual marker genes and genome-wide similarity comparisons revealed that
those infecting AOA are highly divergent from known viruses. While one *Nitrobacter*
142 MAG was obtained and both AOB and NOB viruses were also predicted, the hybrid
approach developed in this study used both ¹²C- and ¹³C-enriched DNA to specifically
144 target low GC mol% DNA and was tailored for the analysis of AOA populations and
their viruses. A broader analysis of all viruses infecting nitrifying prokaryotic
146 populations would therefore likely require a more comprehensive sequencing effort of
all fractions throughout the CsCl gradient using quantitative SIP metagenomics [20] or
148 alternative approaches, such as virus-targeted metagenomic viromes.

150 *Data availability*

Metagenome sequence data are deposited in NCBI's GenBank under BioProject
152 accession nos. PRJNA621418-PRJNA621429 (¹³C data) and PRJNA868779 (¹²C
data).

154

Acknowledgements

156 This work was funded by the AXA Research Fund awarded to GN, ANR grant
'CONSERVE' awarded to CH and GWN, and an EC MSCA Fellowship 'DIVOBIS'
158 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,
160 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
162 Dr. Willm Martens-Habbena (University of Florida) for continued discussions on soil
AOA ecophysiology and taxonomy, Dr. Melina Kerou (University of Vienna) for helpful

164 guidance on AOA MCO diversity and Dr Robin Walker (SRUC, Aberdeen) for access
to the much lamented Craibstone Estate Woodlands Field long term experiment.

166

References

- 168 1. 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.
170 ISME Communications. 2021;1:19.
2. Hink L, Gubry-Rangin C, Nicol GW, Prosser JI. The consequences of niche and
172 physiological differentiation of archaeal and bacterial ammonia oxidisers for
nitrous oxide emissions. ISME J. 2018;12:1084–1093.
- 174 3. Li Y, Chapman SJ, Nicol GW, Yao H. Nitrification and nitrifiers in acidic soils. Soil
Biol Biochem. 2018;116:290–301.
- 176 4. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel,
widespread, and ecologically distinct marine Thaumarchaeota viruses that encode
178 amoC nitrification genes. ISME J. 2019;13:618–631.
5. Kim J-G, Kim S-J, Cvirkaite-Krupovic V, Yu W-J, Gwak J-H, López-Pérez M, et al.
180 Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. Proc
Natl Acad Sci U S A. 2019;116:15645–15650.
- 182 6. 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
184 cohesive responses to moisture in grasslands. bioRxiv. 2022;2022.03.24.485562
- 186 7. Wu R, Davison MR, Gao Y, Nicora CD, Mcdermott JE, Burnum-Johnson KE, et
al. Moisture modulates soil reservoirs of active DNA and RNA viruses. Commun
Biol. 2021;4:992.

- 188 8. Braga LPP, Spor A, Kot W, Breuil M-C, Hansen LH, Setubal JC, et al. Impact of
190 phages on soil bacterial communities and nitrogen availability under different
assembly scenarios. *Microbiome*. 2020;8:52.
- 192 9. Albright MBN, Gallegos-Graves LV, Feeser KL, Montoya K, Emerson JB, Shakya
M, et al. Experimental evidence for the impact of soil viruses on carbon cycling
during surface plant litter decomposition. *ISME Communications*. 2022;2:1–8.
- 194 10. Starr EP, Shi S, Blazewicz SJ, Koch BJ, Probst AJ, Hungate BA, et al. Stable-
Isotope-Informed, Genome-Resolved Metagenomics Uncovers Potential Cross-
196 Kingdom Interactions in Rhizosphere Soil. *mSphere*. 2021;6:e0008521.
- 198 11. Trubl G, Kimbrel JA, Liquet-Gonzalez J, Nuccio EE, Weber PK, Pett-Ridge J, et
al. Active virus-host interactions at sub-freezing temperatures in Arctic peat soil.
Microbiome. 2021;9:208.
- 200 12. 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
202 in soil. *Proc Natl Acad Sci U S A*. 2021;118:e2105124118.
- 204 13. Nicol GW, Leininger S, Schleper C, Prosser JI. The influence of soil pH on the
diversity, abundance and transcriptional activity of ammonia oxidizing archaea
and bacteria. *Environ Microbiol* 2008;10:2966–2978.
- 206 14. 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–
208 1927.
- 210 15. 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 Communications*. 2018;9:1517.

- 212 16. Cardinale DJ, Duffy S. Single-stranded genomic architecture constrains optimal
codon usage. *Bacteriophage*. 2011;1:219–224.
- 214 17. 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*
216 *Biochem*. 2022;166:108569.
18. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, et al. Proteomics
218 and comparative genomics of *Nitrososphaera viennensis* reveal the core genome
and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci U S A*.
220 2016;113:7937–7946.
19. Reyes C, Hodgskiss LH, Kerou M, Pribasnig T, Abby SS, Bayer B, et al. Genome
222 wide transcriptomic analysis of the soil ammonia oxidizing archaeon
Nitrososphaera viennensis upon exposure to copper limitation. *ISME J*.
224 2020;14:2659–2674.
20. Sieradzki ET, Greenlon A, Nicolas AM, Firestone MK, Pett-Ridge J, Blazewicz SJ,
226 et al. Functional succession of actively growing soil microorganisms during
rewetting is shaped by precipitation history. *bioRxiv*. 2022;2022.06.28.498032
228

Figure legends

230 **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 ¹²C- and ¹³C-enriched DNA. **A**
232 Distribution of total prokaryotic and AOA genomes in CsCl gradients determined from
the relative abundance of 16S rRNA genes and *amoA* genes, respectively. Example
234 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
236 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
238 representing an individual microcosm. Genomic DNA in six fractions highlighted in
blue or pink areas were pooled for each replicate microcosm for metagenomic
240 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
242 binned into ten medium- and high-quality nitrifier MAGs and predicted viruses (≥ 10 kb)
of nitrifiers are highlighted, with number in parenthesis beside each MAG identifier
244 giving estimated completeness and contamination (%). **C** Mean and standard error of
the relative proportion of metagenomic reads from ¹²C- and ¹³C-enriched HBD DNA
246 mapped onto contigs of nine AOA and one *Nitrobacter* MAGs. A significantly greater
relative proportion of reads in ¹³C-derived libraries are indicated with * ($p < 0.05$, two-
248 sample Student's t-test or Welsch's t-test when variances were not homogenous).
GTDB genus [14] and *amoA* gene lineage-affiliation [15] are given.

250

Figure 2. Identification and determination of activity of viruses infecting AOA in pH 4.5
252 and 7.5 soils using a hybrid analysis of ¹²C- and ¹³C-enriched DNA. **A** Heatmap
showing the relative abundance of 64 vOTUs identified in LBD DNA libraries from pH

254 4.5 and 7.5 soil ($\geq 1\times$ coverage, $\geq 75\%$ contig breadth). To identify activity, reads from
12C- and 13C-enriched HBD DNA were mapped onto LBD-derived contigs with a
256 significantly greater relative proportion of reads in 13C-derived libraries indicated with
* ($p < 0.05$, two-sample Student's t-test or Welsch's t-test when variances were not
258 homogenous). **B** Proteomic tree showing genome-wide sequence similarities between
AOA viral contigs (VCs) and curated virus genomes. Values at dotted circles represent
260 a distance metric based on normalized tBLASTx scores plotted on a log scale. **C**
Maximum likelihood phylogenetic analysis of derived large sub-unit terminase (TerL)
262 protein sequences (336 unambiguously aligned positions, LG substitution model)
identified in this and other metagenomic studies or in prophages of cultivated AOA.
264 NCBI accession numbers are given in parenthesis. Circles at nodes represent
percentage bootstrap support from 1000 replicates, scale bar denotes an estimated
266 0.05 changes per position and colour-coding describes the comparative relative
abundance in each soil. **D** Maximum likelihood analysis showing phylogenetic
268 placement of a virus genome-encoded type 1 multicopper oxidase (MCO) gene within
the AOA MCO/NirK family of cultivated AOA following designations of Kerou *et al.* [18]
270 (206 unambiguously aligned positions, LG substitution model, invariant and gamma
distributed sites). NCBI accession numbers are given in parenthesis. Circles at nodes
272 represent percentage bootstrap support from 1000 replicates and the scale bar
denotes an estimated 0.05 changes per position.

274

Supplementary Figure 1. Nitrification kinetics in soil microcosms. **A** Ammonium and
276 **B** Nitrite+nitrate concentrations in pH 4.5 and 7.5 soil microcosms after addition of 2 x
100 μg urea-N g^{-1} soil. Ammonia concentrations were determined prior to the addition
278 of urea on day 0 and 15. Points and error bars represent the mean and standard error

value from three destructively sampled microcosms, with some error bars smaller than
280 the symbol. Microcosms were incubated with ^{12}C - CO_2 for determining nitrification
kinetics with no significant differences observed with N concentrations in ^{13}C
282 microcosms at day 30 (data not shown).

284 **Supplementary Figure 2.** Isopycnic centrifugation and metagenomic sequencing of
DNA in targeted fractions for hybrid analysis of ^{12}C - and ^{13}C -enriched DNA. **A**
286 Schematic of GC mol% fractionation in CsCl and selection of samples for
metagenomic sequencing and analysis. **B** Distribution of total prokaryotic, AOA and
288 AOB genomes in CsCl gradients determined from the relative abundance of 16S rRNA
genes (prokaryotes) and *amoA* genes (AOA and AOB). Vertical error bars are the
290 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
292 fractions from three independent CsCl gradients, each representing an individual
microcosm. Fractions highlighted in blue or pink areas were pooled for each replicate
294 microcosm for sequencing. The GC mol% of genomic DNA is given (red text and
arrow) for the ^{12}C fraction containing the highest quantity of genomic DNA for each
296 target group.

298 **Supplementary Figure 3.** Taxonomic affiliation (GTDB) of contigs from prokaryote
genomes (host contigs) or predicted hosts of viruses (virus contigs) in low buoyant
300 density (LBD) and high buoyant density (HBD) metagenomic libraries. AOA of the
class *Nitrososphaeria* are placed within the *Thermoproteota* (GTDB) or
302 *Thaumarchaeota* (NCBI) phylum. Data are the mean values from both pH 4.5 and 7.5
libraries.

304

Supplementary Figure 4. Maximum likelihood phylogenomic analysis of nine AOA

306 MAGs within the class *Nitrososphaeria*. MAGs were compared with cultivated AOA

genomes or reference MAGs using 14,524 aligned positions from 76 single copy

308 genes and rooted with *Nitrosocaldus* representatives. NCBI accession numbers are

given in parenthesis. Circles at nodes represent >95% percentage bootstrap support

310 from 1000 replicates and the scale bar denotes an estimated 0.01 changes per

position.

312

Supplementary Figure 5. Venn diagrams comparing the number of predicted virus

314 contigs using VirSorter1, VirSorter2 and DeepVirFinder tools for pH 4.5 and 7.5 soils.

316 **Supplementary text**

318 **Materials and methods**

Soil microcosms, DNA extraction and stable-isotope probing

320 Soil microcosms were established in 144 ml serum vial bottles with 10 g soil (dry weight equivalent). Microcosms were amended urea solution (100 μg urea-N g^{-1} soil (dry weight)) at day 0 and 15 with the water content being 30% and 32% (w/w) after amendment, respectively. Headspace gas was amended with 5% (v/v) ^{12}C - CO_2 (Air

322 (dry weight)) at day 0 and 15 with the water content being 30% and 32% (w/w) after amendment, respectively. Headspace gas was amended with 5% (v/v) ^{12}C - CO_2 (Air

324 Liquide) or ^{13}C - CO_2 (Sigma-Aldrich) (99% atom enriched) with microcosms opened every 3-4 days to maintain aerobic conditions before re-establishing CO_2

326 concentrations. Microcosms were destructively sampled at day 0, 15 and 30 in triplicate for each treatment and soil stored at -20°C . Ammonium, nitrite and nitrate

328 concentrations were determined immediately after sampling using standard colorimetric assays [1]. Total genomic DNA was extracted from 0.5 g soil using a

330 CTAB-buffer phenol:chloroform:isoamyl alcohol protocol [2] and subject to isopycnic centrifugation in CsCl gradients as previously described (Lee *et al.*, 2021). Briefly, 8

332 ml polyallomer tubes were filled with CsCl dissolved in Tris EDTA buffer (buoyant density 1.71 g ml^{-1}) and 6 μg genomic DNA before sealing and centrifugation at

334 $152,000 \times g$ in an MLN80 rotor (Beckman-Coulter). CsCl gradients were then fractionated into $\sim 350 \mu\text{l}$ aliquots before determining buoyant density via refractive

336 index and recovery of DNA by PEG precipitation [2]. The predicted GC mol% of DNA in specific fractions was determined using the calculation of Schildkraut *et al.* [3].

338

Quantitative PCR and metagenomic sequencing of fractionated DNA samples

340 Quantitative PCR (qPCR) was performed to determine the distribution of prokaryotic
and nitrifier genomes through CsCl gradients using a Corbett Rotor-Gene 6000 real-
342 time PCR cycler. Prokaryotic 16S rRNA, bacterial and archaeal *amoA* genes were
quantified across the entire buoyant density gradients using primer pairs
344 P1(341f)/P2(534r) [4], *amoA1F/amoA2R* [5], and *crenamoA23F/crenamoA616R* [6]),
respectively. Each 25 μ l PCR contained 12.5 μ l 2X QuantiFast SYBR Green Mix
346 (Qiagen), 1 μ M of each primer, 2 μ l of standard DNA, recovered DNA from each
fraction (1/10 dilution) or water (no template negative control). Thermal cycling
348 programs consisted of an initial denaturation step of 15 min at 95°C, followed by 30
cycles of 15 s at 94°C, 30 s at 60°C, 30 s at 72°C for the 16S rRNA gene assay or 15
350 s at 94°C, 30 s at 55°C, 30 s at 72°C for bacterial and archaeal *amoA* genes assays
followed by melt-curve analysis. All assays had an efficiency between 96 and 99%
352 with an r^2 value ≥ 0.99 . After assessing the distribution of prokaryotic and nitrifier
genomes, DNA in fractions with a buoyant density between 1.676 to 1.699 g ml⁻¹ (low
354 buoyant density) and 1.719 to 1.750 g ml⁻¹ (high buoyant density) were pooled for
metagenome sequencing. LBD and HBD metagenomes were sequenced at
356 IntegraGen (Paris, France) and the Joint Genome Institute (JGI, Berkeley, CA, USA),
respectively, both using the NovaSeq sequencing platform (Illumina) with NovaSeq
358 XP version 1 reagent kits and a S4 flowcell 150 bp paired reads.

360 *Assembly, binning, and annotation*

Individual reads of HBD metagenomes were quality-trimmed with the MetaWrap
362 read_qc module [7]. *De novo* co-assembly of the 135 to 245 million quality-controlled
reads per metagenome was performed using MEGAHIT version 1.2.9 [8]. Binning was
364 performed using metaBAT 2 [9], MaxBin2 [10] and CONCOCT [11] implemented in

MetaWRAP version 1.2.1 [7] and bin completion and contamination estimated using
366 CheckM version 1.0.12 [12]. Taxonomic annotation of contigs was performed using
Kaiju [13] with the NCBI nr database (2021-02-24). Taxonomy of MAGs was assigned
368 using GTDB-Tk version 0.3.2 [14] with the Genome Taxonomy Database (GTDB,
release date 2022-03-23).

370

Identification of putative viruses infecting nitrifiers

372 Viral contigs (VCs) were predicted from contigs ≥ 10 kb using VirSorter [15], VirSorter
2.0 [16] and DeepVirFinder [17] (see Supplementary Figure 5 for comparison of
374 predictions). CheckV [18] and manual curation (i.e. identification of viral hallmark
genes such as major capsid, portal, terminase and integrase proteins, enrichment in
376 non-annotated genes) were performed to confirm a potential viral origin of each VC.
The relative abundance of each VC in HBD and LBD metagenomes was determined
378 using the MetaWRAP-Quant-bins module [7] and BMap [19] was used to confirm a
detection threshold of $\geq 75\%$ of contig length with $\geq 1x$ read coverage recruited at $\geq 90\%$
380 average nucleotide identity [20]. The relative abundance of each VC in each HBD and
LBD metagenome was calculated based on the length of sequence size and coverage
382 and abundance was expressed as normalized copies per million reads (CPM). Gene
prediction of each VC was performed using Prodigal version 2.6.3 with meta option
384 [21], and homology search was performed against the NCBI nr database using
Diamond blastp [22] and Interproscan5 [23].

386 To identify contigs from viruses infecting nitrifiers specifically, a database of
virus genes present in proviruses of GTDB nitrifier reference genomes was
388 constructed using VIBRANT [24], PhageBoost [25] and VirSorter [15]. Major capsid,
terminase, portal and integrase protein sequences were then used as search queries

390 against each VC using Diamond blastp (e-value < 10⁻⁵). Genes representing potential
nitrifier-specific auxiliary metabolic genes were manually checked in annotation lists.
392 A homologue-based approach was also used where an AOA host was predicted when
'best hit' shared homologues with those in genomes of *Thaumarchaeota* (NCBI
394 taxonomy) were ≥3x more abundant compared to the second most dominant phylum
[26]. Unlike our previous DNA-SIP analysis of methanotroph populations and
396 associated viruses in these soils [27], spacer sequences in AOA MAG CRISPR arrays
only matched predicted AOA viruses from other studies with 2 mismatches between
398 spacer and protospacer sequences and none from this study (data not shown).

400 *Phylogenomic, phylogenetic and predicted protein analyses*

Phylogenomic analysis of AOA MAGs and reference genomes from cultivated AOA
402 and MAGs was performed using an alignment of single copy marker genes generated
with GToTree [28] with a subsequent maximum likelihood phylogeny calculated using
404 FastTree [29]. Single gene phylogenetic analyses were performed using alignments
generated using MUSCLE [30] and manually refined. Maximum likelihood
406 phylogenetic trees were constructed using PhyML [31] with automatic model selection.
Affiliation of AOA MAGs to *amoA* gene-defined lineages was determined from
408 phylogenetic analysis using the curated reference database of Alves *et al.* [32]. Where
amoA genes were absent in incomplete *Nitrosotalea* MAGs, a broad-level affiliation
410 was inferred from close phylogenomic relationships to complete genomes and their
amoA lineage designation. Comparison of genome-wide similarity between AOA virus
412 contigs and curated virus reference sequences was performed using ViPTree [33] and
the Virus-Host DB [34] to generate a protein tree based on normalised tBLASTx
414 scores.

Supplementary text references

- 416 1. Hink L, Gubry-Rangin C, Nicol GW, Prosser JI. The consequences of niche and
physiological differentiation of archaeal and bacterial ammonia oxidisers for
418 nitrous oxide emissions. *ISME J.* 2018; 12: 1084–1093.
2. Nicol GW, Prosser JI. Strategies to determine diversity, growth, and activity of
420 ammonia-oxidizing archaea in soil. *Methods Enzymol* 2011;496:3–34.
3. Schildkraut CL, Marmur J, Doty P. Determination of the base composition of
422 deoxyribonucleic acid from its buoyant density in CsCl. *J Mol Biol.* 1962;4:430–
443.
- 424 4. Muyzer G, de Waal EC, Uitterlinden AG. Profiling of complex microbial populations
by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-
426 amplified genes coding for 16S rRNA. *Appl Environ Microbiol.* 1993;59:695–700.
5. Rotthauwe JH, Witzel KP, Liesack W. The ammonia monooxygenase structural
428 gene *amoA* as a functional marker: molecular fine-scale analysis of natural
ammonia-oxidizing populations. *Appl Environ Microbiol.* 1997;63:4704–4712.
- 430 6. Tourna M, Freitag TE, Nicol GW, Prosser JI. Growth, activity and temperature
responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ*
432 *Microbiol.* 2008;10:1357–1364.
7. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-
434 resolved metagenomic data analysis. *Microbiome.* 2018;6:1–13.
8. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node
436 solution for large and complex metagenomics assembly via succinct de Bruijn
graph. *Bioinformatics.* 2015;31:1674–1676.

- 438 9. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive
binning algorithm for robust and efficient genome reconstruction from
440 metagenome assemblies. *PeerJ*. 2019;7:e7359.
10. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm
442 to recover genomes from multiple metagenomic datasets. *Bioinformatics*.
2016;32:605–607.
- 444 11. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
metagenomic contigs by coverage and composition. *Nat Methods*. 2014;11:1144–
446 1146.
12. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM:
448 assessing the quality of microbial genomes recovered from isolates, single cells,
and metagenomes. *Genome Res*. 2015;25:1043–1055.
- 450 13. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for
metagenomics with Kaiju. *Nat Commun*. 2016;7: 11257.
- 452 14. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;36: 1925–
454 1927.
15. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from
456 microbial genomic data. *PeerJ*. 2015;3:e985.
16. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al.
458 VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and
RNA viruses. *Microbiome*. 2021;9:37.
- 460 17. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, et al. Identifying viruses
from metagenomic data using deep learning. *Quant Biol*. 2020;8:64–77.

- 462 18. Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosh E, Roux S, Kyrpides NC.
CheckV assesses the quality and completeness of metagenome-assembled viral
464 genomes. *Nat Biotechnol.* 2021;39:578–585.
19. Bushnell B. BBMap: A fast, accurate, splice-aware aligner. 2014. Lawrence
466 Berkeley National Lab. (LBNL), Berkeley, CA (United States).
20. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et
468 al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat
Biotechnol.* 2019;37:29–37.
- 470 21. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
prokaryotic gene recognition and translation initiation site identification. *BMC
472 Bioinformatics.* 2010;11:119.
22. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using
474 DIAMOND. *Nat Methods.* 2015;12:59–60.
23. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:
476 genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–1240.
24. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and
478 curation of microbial viruses, and evaluation of viral community function from
genomic sequences. *Microbiome.* 2020;8:90.
- 480 25. Sirén K, Millard A, Petersen B, Gilbert MTP, Clokie MRJ, Sicheritz-Pontén T.
Rapid discovery of novel prophages using biological feature engineering and
482 machine learning. *NAR Genom Bioinform.* 2021;3:lqaa109.
26. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, et al. Clades of
484 huge phages from across Earth’s ecosystems. *Nature.* 2020;578:425–431.

27. Lee S, Sieradzki ET, Nicolas AM, Walker RL, Firestone MK, Hazard C, et al.
486 Methane-derived carbon flows into host–virus networks at different trophic levels
in soil. *Proc Natl Acad Sci U S A*. 2021; 118: e2105124118.
- 488 28. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics*.
2019;35:4162–4164.
- 490 29. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood
trees for large alignments. *PLoS One*. 2010;5:e9490.
- 492 30. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
throughput. *Nucleic Acids Res*. 2004;32:1792–1797.
- 494 31. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New
algorithms and methods to estimate maximum-likelihood phylogenies: assessing
496 the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–321.
32. Alves RJE, Eloy Alves RJ, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying
498 the global phylogeny and environmental distribution of ammonia-oxidising
archaea based on amoA genes. *Nat Commun*. 2018;9:1517.
- 500 33. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al.
Linking Virus Genomes with Host Taxonomy. *Viruses*. 2016;8:66.
- 502 34. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the
viral proteomic tree server. *Bioinformatics*. 2017;33:2379–2380.
- 504

Figure 1

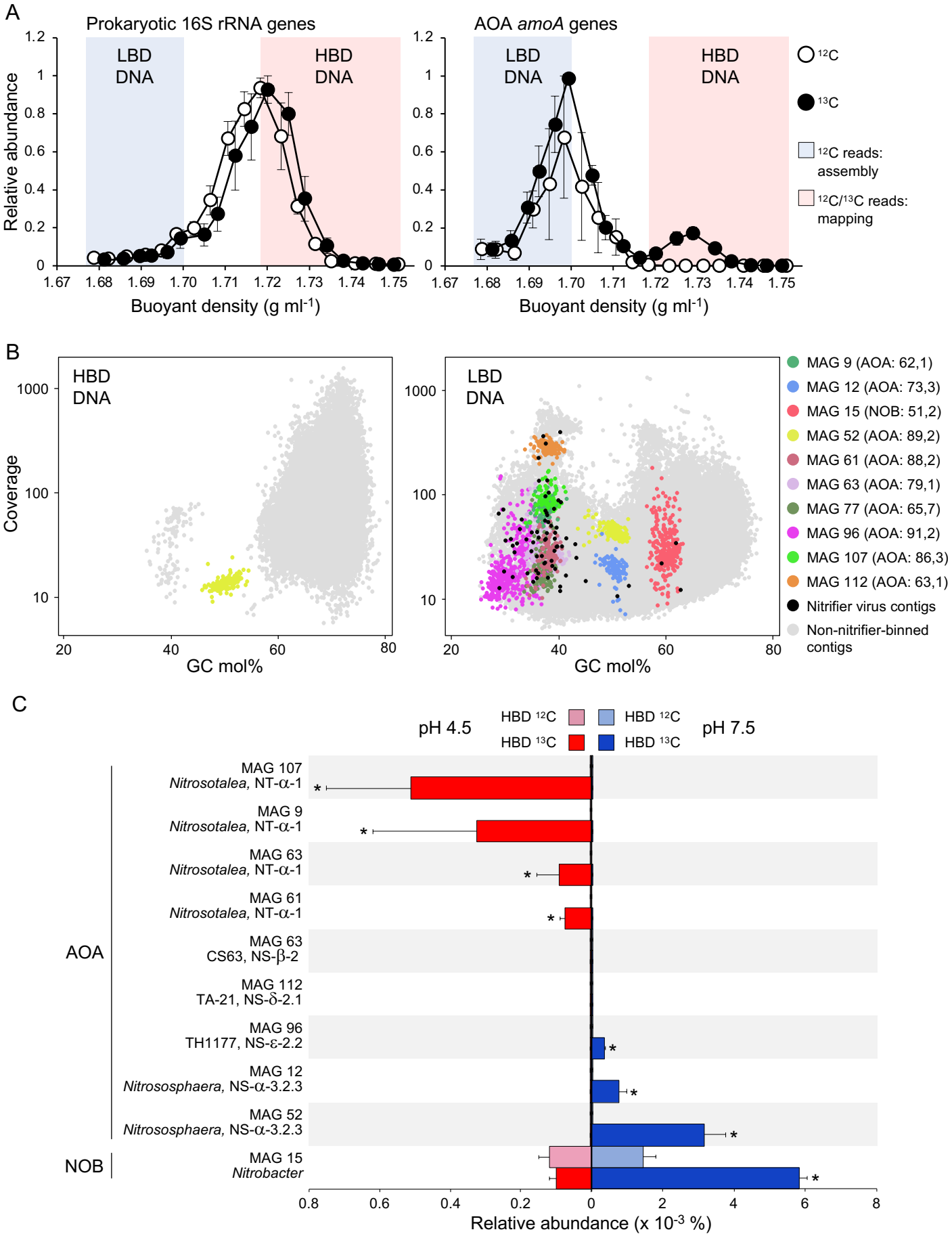
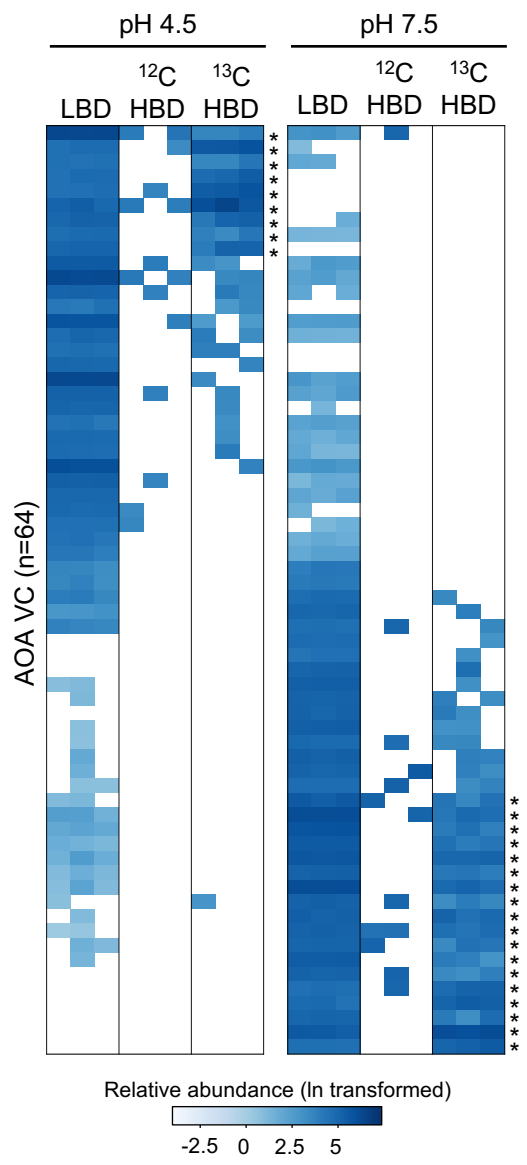
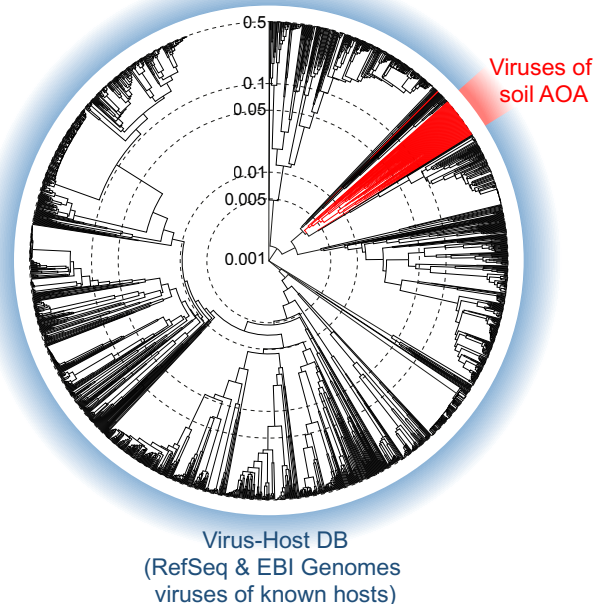


Figure 2

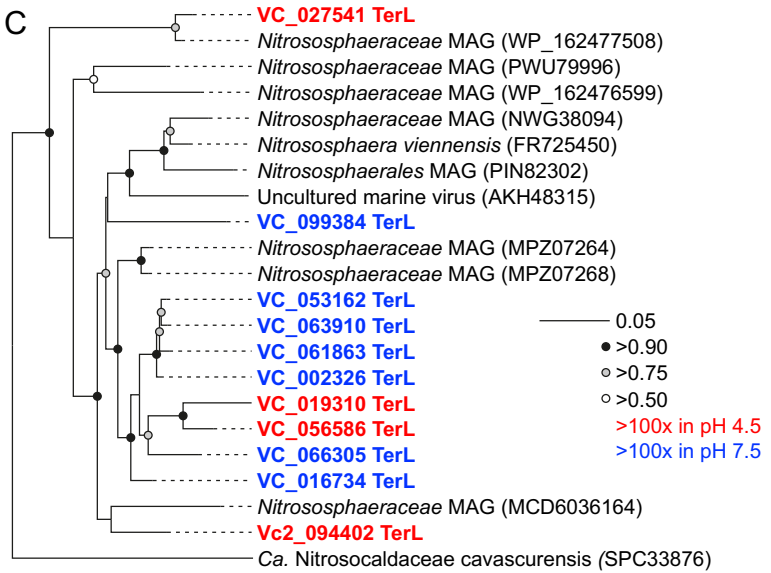
A



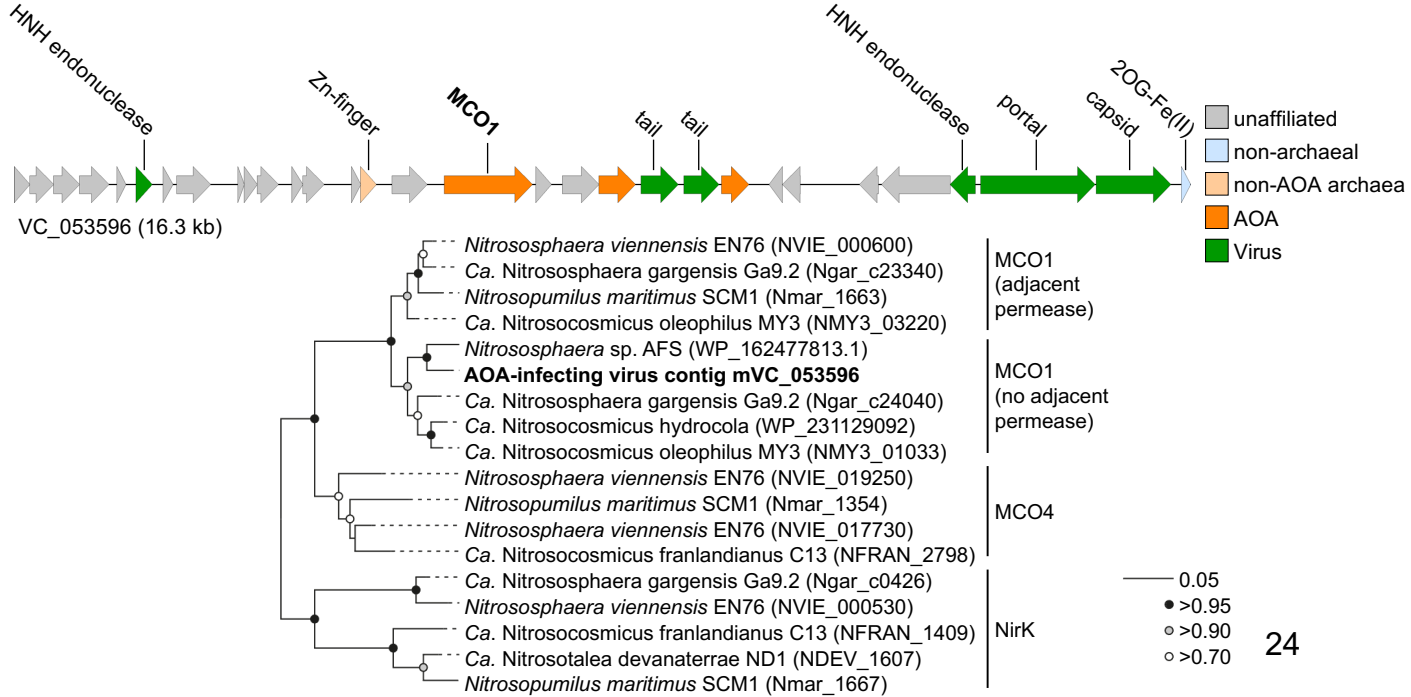
B

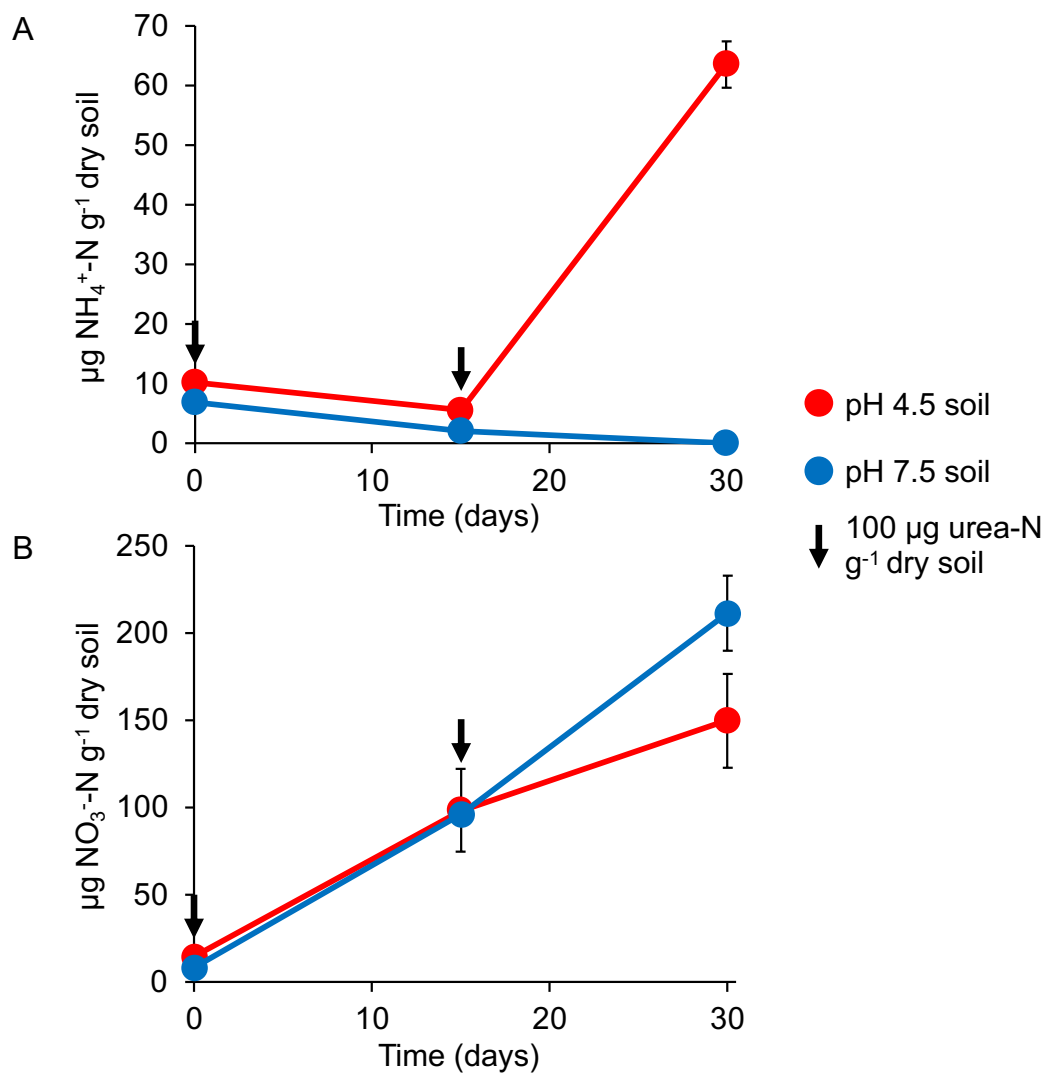


C

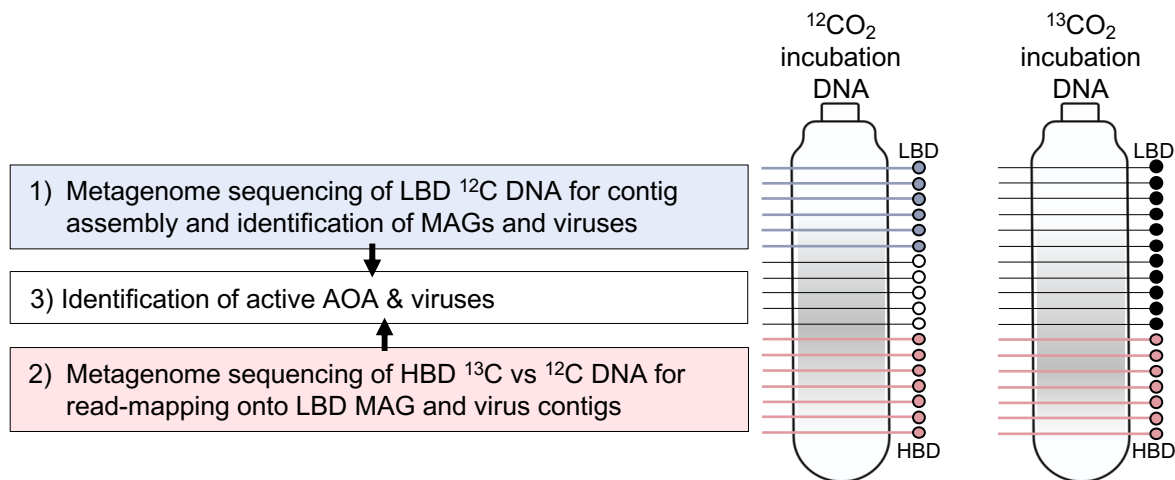


D

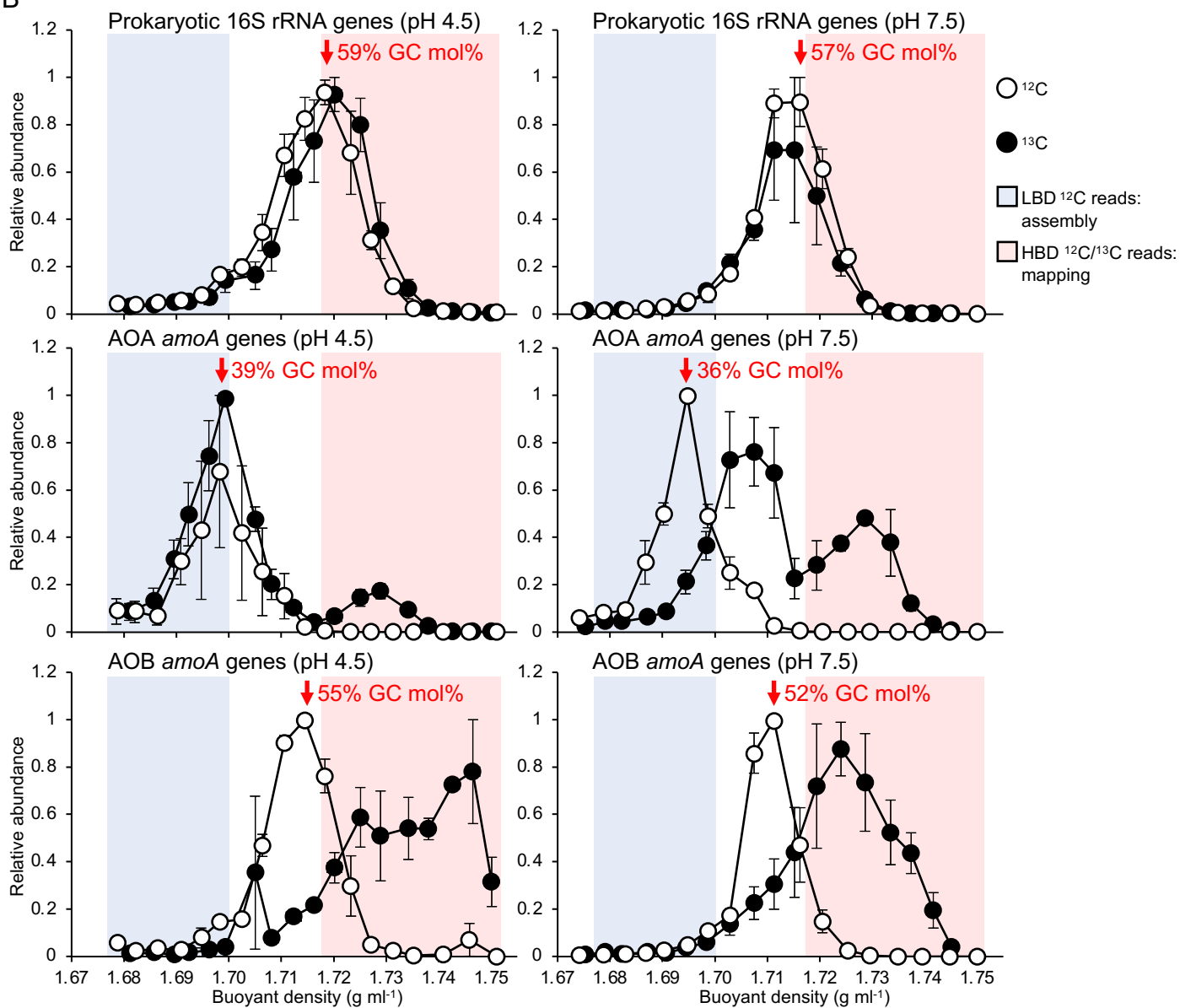




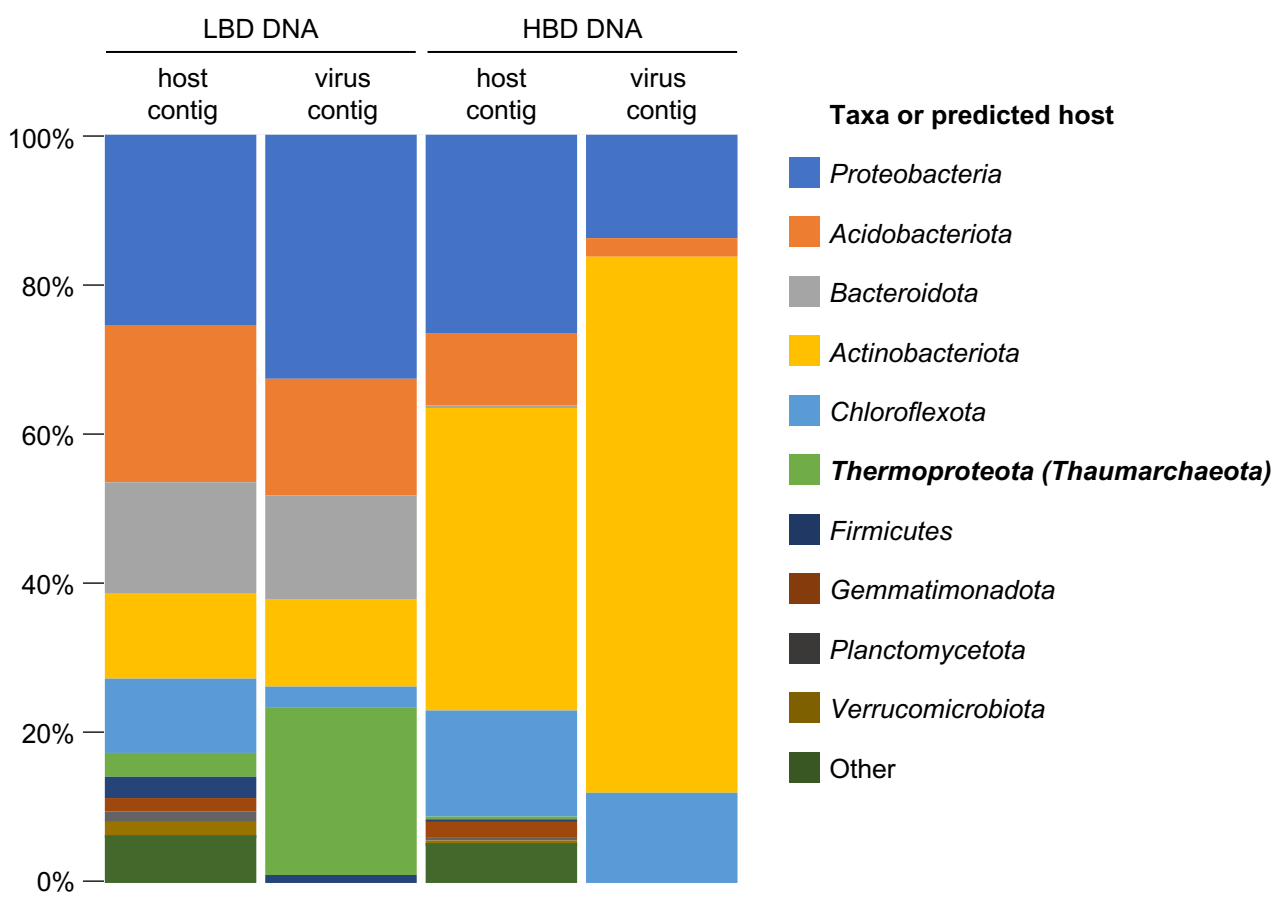
A

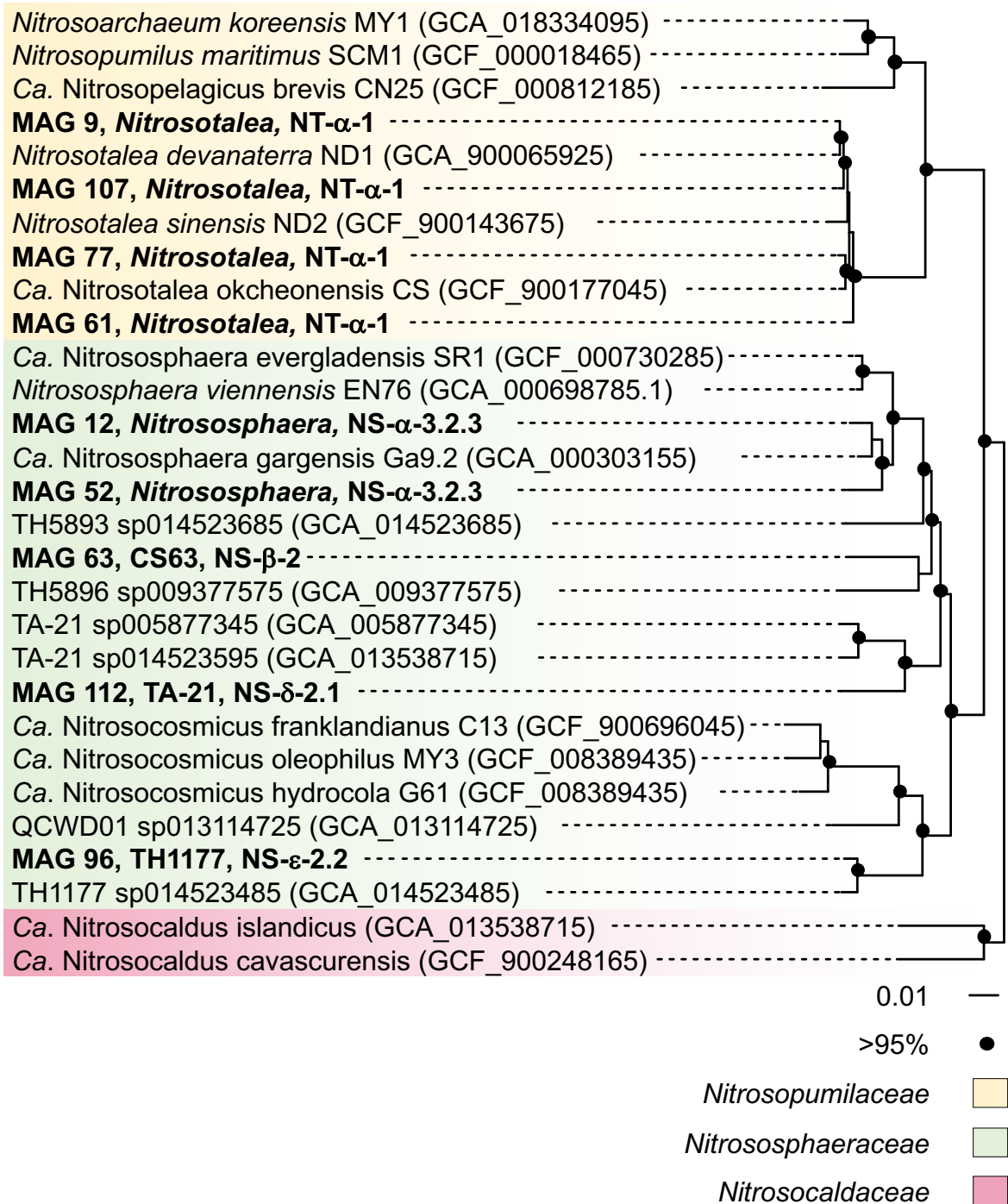


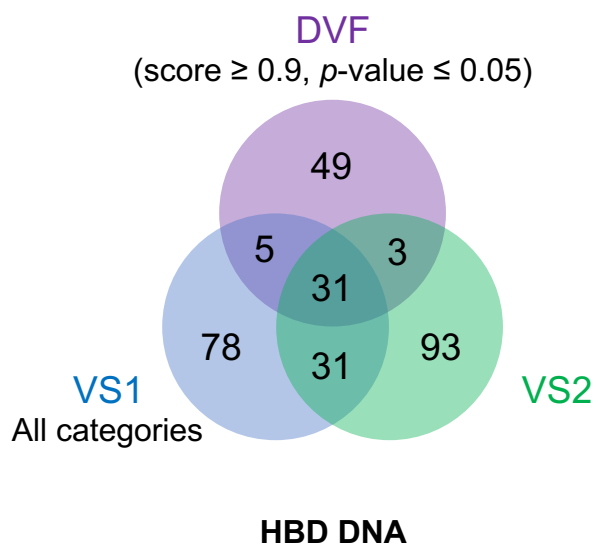
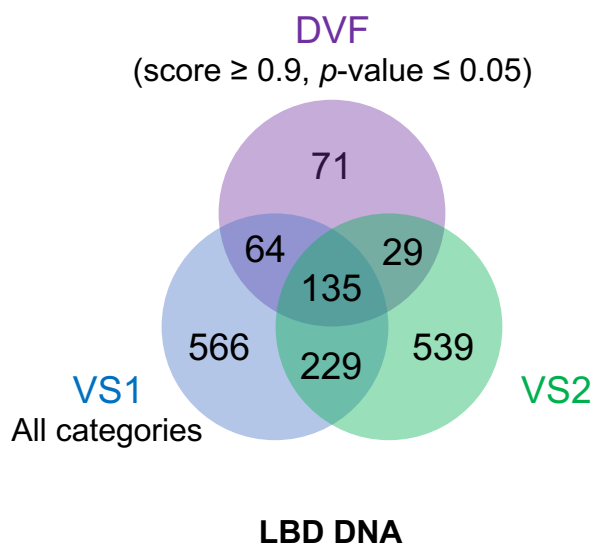
B



Supplementary Figure 3







Supplementary Table 1. Summary of metagenome sequencing of genomic DNA from low buoyant density (LBD) and high buoyant density (HBD) CsCl fractions derived from pH 4.5 and 7.5 soil. **A** Library accession numbers and read depth. **B** Contig assembly details.

A

Sample ID*	NCBI BioProject Accession	Post-QC number of reads
LBD-12C-pH45_1	PRJNA868779	170,519,430
LBD-12C-pH45_2	PRJNA868779	190,442,846
LBD-12C-pH45_3	PRJNA868779	183,142,278
LBD-12C-pH75_1	PRJNA868779	178,892,472
LBD-12C-pH75_2	PRJNA868779	180,678,072
LBD-12C-pH75_3	PRJNA868779	165,641,626
HBD-12C-pH45_1	PRJNA621418	137,927,826
HBD-12C-pH45_2	PRJNA621419	165,499,900
HBD-12C-pH45_3	PRJNA621420	161,786,148
HBD-13C-pH45_1	PRJNA621421	157,372,294
HBD-13C-pH45_2	PRJNA621422	140,157,446
HBD-13C-pH45_3	PRJNA621423	153,263,184
HBD-12C-pH75_1	PRJNA621424	172,033,988
HBD-12C-pH75_2	PRJNA621425	184,944,656
HBD-12C-pH75_3	PRJNA621426	173,060,618
HBD-13C-pH75_1	PRJNA621427	135,351,318
HBD-13C-pH75_2	PRJNA621428	245,188,756
HBD-13C-pH75_3	PRJNA621429	156,686,154

*name describes HBD or LBD CsCl fractions, ¹²C or ¹³C incubation, soil pH and replicate number.

B

Sample ID	Assembled contigs	Mean contig length (bp)	Max. contig length (bp)
LBD-co-assembled contig	26,611,526	614.2	809,772
5kb-LBD-co-assembled contig	117,750	9,437.5	
HBD-co-assembled contig	35,109,680	566.3	812,893
5kb-HBD-co-assembled contig	76,311	8,810.5	

Supplementary Table 2. Summary of 123 medium- and high-quality metagenome assembled genomes from low buoyant density DNA ranked by phylum (GTDB taxonomy).

MAG ID	Completeness (%)	Contamination (%)	GC mol%	N50 (bp)	Size (bp)	Phylum	Class	Order	Family	Genus
75	83.7	0.2	66.9	24,140	2,864,512	Acidobacteriota	Acidobacteriae	Acidobacteriales		
56	82.5	0.0	60.3	15,391	3,027,569	Acidobacteriota	Acidobacteriae	Acidobacteriales	Acidobacteriaceae	PALSA-350
101	80.7	5.1	54.0	21,626	3,847,244	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae	Gp1-AA145
126	77.1	3.0	67.5	17,321	3,322,481	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999	
72	73.9	0.9	58.4	14,749	2,220,118	Acidobacteriota	Acidobacteriae	UBA7541	UBA7541	
124	67.3	3.4	63.9	11,528	2,512,459	Acidobacteriota	Acidobacteriae	Acidobacteriales		
33	66.0	5.5	59.3	15,405	2,887,037	Acidobacteriota	Acidobacteriae	UBA7541	UBA7541	Palsa-295
22	53.3	0.9	56.3	9,893	2,452,695	Acidobacteriota	Acidobacteriae	UBA7541	UBA7541	
68	50.8	3.4	61.9	9,900	2,500,689	Acidobacteriota	Acidobacteriae	Acidobacteriales	Acidobacteriaceae	KBS-83
87	50.4	2.0	52.4	8,779	1,812,072	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	OLB17
111	97.3	1.9	66.6	36,117	2,871,784	Actinobacteriota	Thermoleophila	Solirubrobacterales	Solirubrobacteraceae	
44	97.0	2.6	66.0	142,875	3,274,599	Actinobacteriota	Thermoleophila	Solirubrobacterales	Solirubrobacteraceae	
25	97.0	3.1	65.7	71,608	2,181,555	Actinobacteriota	Thermoleophila	Solirubrobacterales	70-9	
46	82.0	0.1	63.9	17,436	1,359,101	Actinobacteriota	Acidimicrobiia	Acidimicrobiales	RAAP-2	RAAP-2
6	79.2	2.2	68.5	54,707	2,098,665	Actinobacteriota	Thermoleophila	Solirubrobacterales	70-9	
57	70.8	1.2	69.6	12,884	2,594,437	Actinobacteriota	Actinobacteria	Nanopelagiales		
48	63.0	3.0	70.1	11,121	2,066,628	Actinobacteriota	Acidimicrobiia	Acidimicrobiales	RAAP-2	Bog-756
93	60.3	0.9	62.8	10,343	1,801,421	Actinobacteriota	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	UBA668
79	67.5	0.9	61.7	11,258	2,006,440	Armatimonadota	Fimbriimonadia	Fimbriimonadales	Fimbriimonadaceae	
38	96.5	0.7	44.9	52,178	3,501,364	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	JJ008
100	94.8	1.5	39.6	298,057	4,276,415	Bacteroidota	Bacteroidia	AKYH767		
53	92.0	1.7	44.7	26,881	3,591,767	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	UBA8621
62	86.9	7.5	38.5	28,454	3,315,570	Bacteroidota	Bacteroidia	AKYH767-A	OLB10	
76	84.0	0.7	49.6	26,115	3,041,112	Bacteroidota	Bacteroidia	AKYH767	2-12-FULL-35-15	
97	78.6	2.1	39.6	15,282	3,408,942	Bacteroidota	Bacteroidia	AKYH767		
81	77.5	0.2	54.5	15,405	2,291,910	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	
106	76.6	3.4	44.4	12,939	2,749,566	Bacteroidota	Bacteroidia	AKYH767	b-17BO	PALSA-968
32	75.2	3.8	40.6	13,274	4,299,582	Bacteroidota	Bacteroidia	AKYH767-A	OLB10	
47	73.6	4.2	34.8	11,826	2,228,512	Bacteroidota	Ignavibacteria	Ignavibacteriales	Ignavibacteriaceae	BMS3ABIN03
116	73.2	1.9	51.5	13,585	2,332,193	Bacteroidota	Kapabacteria	Palsa-1295	Palsa-1295	
13	71.5	0.0	42.1	13,719	2,460,739	Bacteroidota	Bacteroidia	NS11-12g	UBA955	
30	67.2	0.5	37.7	10,543	1,634,027	Bacteroidota	Bacteroidia	Sphingobacteriales	Sphingobacteriaceae	
26	66.5	1.5	42.0	11,489	3,546,402	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	
86	65.7	3.4	34.7	9,330	2,102,202	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavobacterium
89	65.4	0.0	39.1	18,038	2,305,465	Bacteroidota	Ignavibacteria	SJA-28	OLB5	
40	64.5	7.3	39.0	8,876	2,679,712	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	Taibaiella_B
118	52.0	3.2	38.0	9,613	3,001,274	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	Palsa-955
41	51.3	4.4	39.4	8,197	2,857,311	Bacteroidota	Bacteroidia	Chitinophagales	BACL12	UBA7236
83	50.8	1.0	41.9	11,602	2,718,158	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	
49	52.6	1.8	42.5	10,353	1,916,736	Bdellovibrionota	Bdellovibrionia	Bdellovibrionales	Bdellovibrionaceae	Bdellovibrio
58	91.4	0.0	61.2	14,550	3,678,188	Chloroflexota	UBA5177	UBA5177	UBA5177	
69	75.9	2.2	62.3	12,810	2,934,925	Chloroflexota	UBA4733	UBA4733	UBA4733	
115	75.1	2.8	70.3	18,220	2,481,858	Chloroflexota	Ellin6529	CSP1-4	CSP1-4	Palsa-1033
16	64.2	1.4	55.1	23,141	1,640,525	Chloroflexota	Anaerolineae	Anaerolineales	UBA11579	
121	63.5	8.2	48.7	8,645	3,693,765	Chloroflexota	Anaerolineae	Anaerolineales	envOPS12	OLB14
14	62.9	1.0	68.7	11,300	3,380,230	Chloroflexota	Ktedonobacteria	Ktedonobacteriales		

36	59.5	2.8	72.0	10,127	1,826,549	Chloroflexota	Ellin6529	CSP1-4	CSP1-4	CSP1-4
17	57.0	0.9	64.1	9,099	3,136,804	Chloroflexota	UBA5177			
24	56.3	0.9	63.3	7,671	2,328,854	Chloroflexota	UBA5177			
43	92.3	0.9	43.4	23,016	2,895,603	Cyanobacteria	Vampirovibrionia	Vampirovibrionales		
104	52.6	0.0	46.6	9,501	1,686,795	Cyanobacteria	Vampirovibrionia	Vampirovibrionales		
84	62.1	0.0	35.2	9,392	965,279	Dependentiae	Babeliae	Babeliales		
11	60.2	0.9	69.4	14,745	1,971,573	Dormibacterota	Dormibacteria	UBA8260	UBA8260	
23	54.3	9.5	68.9	8,592	1,357,015	Dormibacterota	Dormibacteria	UBA8260	UBA8260	Palsa-851
119	87.2	0.0	38.0	15,691	2,341,565	Eremiobacterota	Eremiobacteria	Eremiobacterales	Eremiobacteraceae	
103	77.4	1.5	68.2	12,265	2,115,044	Eremiobacterota	Eremiobacteria	UBP12	UBA5184	
37	51.1	7.4	60.2	7,077	1,974,572	Eremiobacterota	Eremiobacteria	UBP12	UBA5184	PALSA-1484
82	95.1	0.3	34.9	35,039	2,302,984	Firmicutes	Clostridia	Lachnospirales	Lachnospiraceae	Herbinix
70	84.4	2.6	41.4	17,334	3,226,646	Firmicutes	Bacilli	Bacillales	Bacillaceae	
10	78.7	0.8	45.6	21,297	2,470,344	Firmicutes	Bacilli	Bacillales	Bacillaceae	
123	59.2	1.0	36.4	18,393	979,880	Firmicutes	Bacilli	Bacillales	Planococcaceae	Ureibacillus
99	54.5	9.9	36.8	6,988	796,782	Firmicutes	Bacilli	Bacillales	Bacillaceae_G	Bacillus
3	51.9	5.3	32.6	12,001	1,498,041	Firmicutes	Clostridia	Acetivibrionales	Acetivibrionaceae	Herbivorax
80	97.4	0.7	41.7	35,221	2,034,588	Firmicutes_F	Halanaerobiia	Halanaerobiales	DTU029	DTU029
35	99.2	1.0	42.0	79,014	2,517,665	Firmicutes_I	Bacilli	Thermoactinomycetales	Thermoactinomycetaceae	
120	96.5	0.3	48.2	35,048	2,270,355	Firmicutes_I	Bacilli	Thermoactinomycetales	Thermoactinomycetaceae	CDF
67	69.4	0.3	41.6	9,016	1,357,832	Firmicutes_I	Bacilli	Thermoactinomycetales	Thermoactinomycetaceae	
122	71.8	2.7	65.3	16,326	2,944,945	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	FEN-1250
34	71.6	3.4	66.1	29,597	3,506,635	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	
39	69.1	1.2	67.2	17,110	3,202,677	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	
91	52.7	0.0	62.8	9,236	2,678,084	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	AG2
117	58.9	0.0	68.6	7,421	2,581,217	Myxococcota	Polyangia	Palsa-1104	Palsa-1104	PALSA-1104
31	78.3	0.0	35.0	63,246	1,110,445	Patescibacteria	Dojkabacteria			
110	72.5	0.0	45.6	538,338	833,363	Patescibacteria	Saccharimonadia	Saccharimonadales	2-12-FULL-41-12	
102	70.0	0.0	35.8	24,450	1,170,104	Patescibacteria	Microgenomatia	Levybacterales	UBA12049	
60	69.9	1.9	63.1	809,772	809,772	Patescibacteria	Saccharimonadia			
51	65.0	0.8	58.3	36,742	983,527	Patescibacteria	Paceibacteria	UBA6257	2-01-FULL-56-20	
50	64.6	0.0	39.7	277,019	821,951	Patescibacteria	Microgenomatia	UBA1406	HO2-37-13b	
66	64.5	0.0	41.9	35,340	846,808	Patescibacteria	Microgenomatia	UBA1406	GWC2-37-13	2-01-FULL-40-42
42	64.4	0.0	43.7	72,802	632,826	Patescibacteria	Saccharimonadia	Saccharimonadales	UBA10212	
7	64.2	0.0	53.8	40,147	643,112	Patescibacteria	Saccharimonadia	Saccharimonadales	UBA10212	
29	60.9	1.0	43.9	26,687	688,051	Patescibacteria	Doudnabacteria	UBA920	O2-02-FULL-48-8	
45	59.9	0.0	39.8	12,821	852,036	Patescibacteria	Microgenomatia	Levybacterales	UBA12049	PRDT01
28	59.4	0.0	49.9	16,270	829,123	Patescibacteria	Saccharimonadia	Saccharimonadales	UBA4665	UBA6224
74	54.6	0.0	35.5	26,797	536,454	Patescibacteria	Microgenomatia	Levybacterales	UBA12049	
114	53.4	6.5	43.9	9,739	841,384	Patescibacteria	Saccharimonadia	Saccharimonadales	2-12-FULL-41-12	
27	52.7	0.4	46.8	40,409	442,116	Patescibacteria	Paceibacteria	UBA9983	Zambryskibacteraceae	UBA5004
5	51.9	0.0	37.5	13,915	425,834	Patescibacteria	Paceibacteria	UBA9983	Zambryskibacteraceae	C7867-006
54	50.4	0.0	38.2	21,102	760,548	Patescibacteria	Microgenomatia	Woykebacterales		
8	66.5	0.0	58.4	10,752	3,564,280	Planctomycetota	Phycisphaerae	UBA1161	UBA1161	
105	65.4	5.8	59.4	11,513	4,165,716	Planctomycetota	Phycisphaerae	UBA1161	UBA1161	
73	95.1	4.3	65.6	137,474	4,398,083	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Steroidobacteraceae	
19	89.5	2.6	70.1	31,514	3,812,969	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	BOG-938
125	88.0	3.1	62.4	28,686	2,191,964	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
108	87.5	6.2	57.0	22,826	1,972,878	Proteobacteria	Alphaproteobacteria	Micavibrionales	Micavibrionaceae	UM-FILTER-47-13
92	82.8	5.0	65.9	19,149	2,660,651	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Steroidobacteraceae	13-2-20CM-66-19
64	81.0	1.7	61.6	22,864	2,553,403	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rudaea
71	79.5	0.9	61.2	22,385	2,981,098	Proteobacteria	Alphaproteobacteria	UBA1301	UBA1301	UBA6038

98	79.3	2.1	70.5	22,416	2,489,191	Proteobacteria	Gammaproteobacteria	Steroidobacteriales	Steroidobacteraceae	13-2-20CM-66-19
94	78.7	0.9	59.3	14,277	2,998,862	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae	Pseudolabrys
95	77.7	0.0	35.2	13,606	3,485,685	Proteobacteria	Gammaproteobacteria	UBA5158	UBA5158	
65	77.3	1.1	66.2	16,642	5,023,705	Proteobacteria	Alphaproteobacteria	Acetobacterales	Acetobacteraceae	Palsa-883
2	73.3	1.4	67.0	20,592	2,309,631	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Porphyrobacter
109	61.9	5.3	63.3	13,874	3,496,047	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae	BOG-931
78	56.7	1.3	64.9	35,205	1,728,632	Proteobacteria	Alphaproteobacteria	Elsterales	URHD0088	
55	56.4	0.8	61.4	11,298	1,368,874	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methyloligellaceae	Methyloceanibacter
90	55.5	2.1	67.2	11,614	1,755,960	Proteobacteria	Gammaproteobacteria	Burkholderiales	Palsa-1005	PALSA-1003
88	55.2	3.4	65.1	9,131	1,518,294	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae	Palsa-892
18	52.5	5.7	59.0	7,355	2,371,465	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methyloligellaceae	Methyloceanibacter
20	51.7	0.0	61.0	9,057	2,689,297	Proteobacteria	Alphaproteobacteria	UBA1301	UBA1301	
15	51.2	1.7	59.9	9,856	2,680,085	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae	Nitrobacter
21	50.3	1.1	64.3	8,593	1,496,296	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	C04
96	91.3	1.9	31.0	10,635	4,039,103	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
52	88.8	1.9	49.7	15,865	1,740,480	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
61	87.9	1.9	37.8	14,936	1,491,779	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotalea
107	85.9	2.9	37.8	36,213	1,658,330	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotalea
63	78.6	1.0	37.4	10,792	2,066,632	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	
12	73.3	3.2	50.2	11,601	1,019,284	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
77	65.2	6.5	36.8	16,902	1,553,150	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotalea
112	62.9	1.0	37.9	9,638	1,212,031	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TA-21
9	61.7	1.0	37.3	13,015	970,273	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotalea
85	98.6	8.6	57.1	36,458	4,399,804	Verrucomicrobiota	Verrucomicrobiae	Pedosphaerales	Pedosphaeraceae	UBA11358
59	94.6	1.6	37.8	26,369	2,935,834	Verrucomicrobiota	Chlamydia	Parachlamydiales	Parachlamydiaceae	