



**HAL**  
open science

## Trace gas oxidizers are widespread and active members of soil microbial communities

Sean Bay, Xiyang Dong, James Bradley, Pok Man Leung, Rhys Grinter, Thanavit Jirapanjawat, Stefan Arndt, Perran Cook, Douglas Larowe, Philipp Nauer, et al.

### ► To cite this version:

Sean Bay, Xiyang Dong, James Bradley, Pok Man Leung, Rhys Grinter, et al.. Trace gas oxidizers are widespread and active members of soil microbial communities. *Nature Microbiology*, 2020, 6 (2), pp.246-256. 10.1038/s41564-020-00811-w . hal-04527857

**HAL Id: hal-04527857**

**<https://hal.science/hal-04527857>**

Submitted on 31 Mar 2024

**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.

1  
2  
3  
4  
5

## 1. Extended Data

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: <i>Smith_ED Fig1.jpg</i>	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	Composition of the bacterial and archaeal communities sequenced in each soil metagenome.	Extended-Data-Figure-1.tif	Stacked barcharts depicting the relative abundance of different phyla in <b>(a)</b> Australian soils and <b>(b)</b> global soils based on reads for the single-copy ribosomal protein gene <i>rplP</i> . Alpha diversity is shown as observed and estimated richness (Chao1) in <b>(c)</b> Australian soils and <b>(d)</b> global soils. Boxplots show median, lower and upper quartile, and minimum and maximum values. This is based on four biological replicates for the four Australian soils (n = 16) at depths of 0-5, 5-10, 15-20, and 25-30 cm, and biological triplicates for the eight global soils (n = 24). For beta diversity, abundance-based distance matrix Bray-Curtis diversity is visualized on a multidimensional PcoA plot for both Australian and global soils. Testing for significant differences in community structure between depths and ecosystems was performed using a one-way PERMANOVA with 999 permutations.
Extended Data Fig. 2	Metatranscriptomic analysis comparing metabolic marker gene expression across three	Extended-Data-Figure-2.tif	<b>(a)</b> Relative abundance of quality-filtered metatranscriptomic short reads expressed as the total reads per kilobase per million (RPKM). <b>(b)</b> Relative abundance of genes in each metatranscriptome mapped to the corresponding assemblies

	global soils with paired metagenome and metatranscriptome datasets.		from each soil expressed as total transcripts per million (TPM).
<b>Extended Data Fig. 3</b>	Analysis of aerobic methane oxidation enzymes and pathways in sampled soils.	Extended-Data-Figure-3.tif	<p><b>(a)</b> Maximum-likelihood tree of amino acid sequences of particulate methane monooxygenase A subunit (PmoA), a marker for aerobic methane oxidation. The tree shows 2 sequences from soil metagenome-assembled genomes (blue) and 12 sequences from unbinned contigs (red) alongside 93 representative sequences from NCBI reference genomes (black). The tree shows the affiliation of the PmoA from <i>Candidatus Methylotheobacterium kingii</i> with those of amplicons of the tropical upland soil cluster (TUSC). Also shown are reference sequences of other members of the copper membrane monooxygenase superfamily, namely ammonia monooxygenase (AmoA), hydrocarbon monooxygenase (HmoA), and groups of unknown function (including PxmA). The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted. Source data is provided in Newick (nwk) format. <b>(b)</b> Metabolic reconstruction of the putative novel methanotroph <i>Candidatus Methylotheobacterium kingii</i>. The core pathways associated with energy conservation and carbon acquisition are shown, with genes detected shown in italics. The bacterium is predicted to use methane, methanol, and acetate as energy and carbon sources. In addition, it can use molecular hydrogen as an electron donor via a group 1f [NiFe]-hydrogenase. The bacterium is predicted to use the electron acceptors oxygen <i>via</i> a cytochrome <i>c</i> oxidase and nitrous oxide <i>via</i> a nitrous oxide reductase. Its particulate methane monooxygenase forms a distinct phylogenetic lineage with amplicons from the Tropical Upland Soil Cluster (TUSC), whereas its methanol dehydrogenase is closely related to those in previously sequenced Gemmatimonadota MAGs inferred to be methylotheobacterial. The genome encodes key</p>

			<p>enzymes for the serine cycle for assimilation of one-carbon sources. Abbreviations: H<sub>4</sub>F = tetrahydrofolate; Hyd = group 1f [NiFe]-hydrogenase; pMMO = particulate methane monooxygenase; MDH = methanol dehydrogenase; PQQ = pyrroloquinoline quinone; I = NADH dehydrogenase (complex I); II = succinate dehydrogenase (complex II); IV = cytochrome <i>aa</i><sub>3</sub> oxidase (complex IV). Dashed black lines indicate diffusion. Dashed gray lines indicate unknown process or undetected genes. <b>(c)</b> Molecular model of the putative particulate methane monooxygenase (Pmo) from <i>Candidatus</i> Methylophilum kingii and comparison with putative catalytic sites of other experimentally validated Pmos. (i) Molecular model of the functional homotrimer of the putative Pmo complex from <i>Ca. M. kingii</i>, shown as a cartoon representation. PmoA, PmoB, and PmoC subunits from one Pmo complex are colored in light blue, dark blue, and sky blue respectively. The other two Pmo complexes in the trimer are colored in transparent yellow. Cu ions bound in the putative active site of the color-coded Pmo complex are shown as ochre spheres. TM = transmembrane. (ii) Zoomed view of the Pmo complex from panel A, with the three Pmo subunits labelled and the putative active sites Cu<sub>B</sub> and Cu<sub>C</sub> highlighted. (iii) Stick representation of residues in the Cu<sub>B</sub> and Cu<sub>C</sub> active sites of Pmo molecular models from <i>Ca. M. kingii</i> (Gemmatimonadota) and <i>Methylophilum inferorum</i> (Verrucomicrobiota), and from the crystal structure (PDB ID: 3RGB) of Pmo from the <i>Methylococcus capsulatus</i> (Proteobacteria), showing that both sites are conserved between Pmo from <i>Ca. M. kingii</i> and <i>M. capsulatus</i>. Pmo from <i>M. inferorum</i> lacks the Cu<sub>B</sub> site, suggesting this site is not responsible for methane oxidation. (iv) Sequence alignment of metal binding motifs from the Cu<sub>B</sub> (left) and Cu<sub>C</sub> (right) sites from Pmo from <i>Ca. M. kingii</i> (<i>cMk.</i>), <i>M. capsulatus</i> (<i>Mc.</i>), <i>Candidatus</i> Methylophilum limnetica (<i>cMl.</i>; Methylophilum limnetica), and <i>M. inferorum</i> (<i>Mi.</i>). Amino acids involved in Cu coordination are highlighted.</p>
--	--	--	---



<p><b>Extended Data Fig. 4</b></p>	<p>Biogeochemical measurements of <i>in situ</i> soil gas concentrations and validation of static chamber setup.</p>	<p>Extended-Data-Figure-4.tif</p>	<p><b>(a)</b> Mean soil-gas profiles normalized to the respective ambient air concentration (dashed line) of four biological replicates collected at depths of 0, 2, 4, 6, 8, 10, and 16 cm across four Australian ecosystems (n = 112). Note that the different gases were sampled at identical depths, but points are plotted slightly offset on the y-axis for better visibility of error bars. <b>(b)</b> Laboratory static chamber incubations (n = 3) to control for abiotic release of trace gases from chamber's plastic components. Chamber headspace mixing ratios of hydrogen (H<sub>2</sub>), carbon monoxide (CO), and methane (CH<sub>4</sub>) were measured at eight time points during ~25 minutes of incubation time of an inert stainless steel surface. Measurement setup, sampling procedure, and sampling frequency followed those applied during the <i>in situ</i> chamber incubations performed to measured soil-atmosphere flux of trace gases. Circles indicate mixing ratio averaged from three independent incubations, with the vertical lines denoting one standard deviation. Dashed lines indicate the mixing ratio of trace gas averaged across three air samples.</p>
<p><b>Extended Data Fig. 5</b></p>	<p>Maximum-likelihood tree of amino acid sequences of group 3 [NiFe]-hydrogenase large subunits, a marker for hydrogen production during fermentation processes.</p>	<p>Extended-Data-Figure-5.tif</p>	<p>The tree shows 129 sequences from soil metagenome-assembled genomes (blue) alongside 172 representative sequences from NCBI reference genomes (black). The subgroup of each reference sequence is denoted according to the HydDB classification scheme<sup>41</sup>. The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted. All sequences shorter than 350 amino acids were omitted. Source data is provided in Newick (nwk) format.</p>
<p><b>Extended Data Fig. 6</b></p>	<p>Gas chromatography studies of one <i>ex situ</i> trace gas consumption experiment.</p>	<p>Extended-Data-Figure-6.tif</p>	<p>Shown are four biological replicates collected at depths of 0-5, 5-10, 15-20, and 25-30 cm across four Australian ecosystems (n = 64). Depicted are the oxidation rates recorded over time of <b>(a)</b> atmospheric H<sub>2</sub>, <b>(b)</b> atmospheric CO, and <b>(c)</b> atmospheric CH<sub>4</sub> by soils at each depth compared to heat-</p>

			<p>killed control soils. Oxidation of the three gases was also measured in the dryland soils following hydration, i.e. Dryland (wet) samples. Points show average values and vertical lines the error bars representing one standard deviations of four biological replicates. Significance testing of differences in <i>ex situ</i> oxidation rates between ecosystem type and soil depth was carried out using a one-sided Kruskal Wallis H test followed by a pairwise Wilcoxon tests with adjusted <i>p</i>-values using a Benjamini Hochberg correction for multiple testing. Boxplots show median, lower and upper quartile, and minimum and maximum values comparing the <i>ex situ</i> rates of H<sub>2</sub>, CO, and CH<sub>4</sub> oxidation between depths and ecosystems for <b>(d)</b> cell-specific rates and <b>(e)</b> bulk reaction rates per gram of dry soil.</p>
<b>Extended Data Fig. 7</b>	Copy number of the 16S rRNA gene per gram of soils of four biological replicates.	Extended-Data-Figure-7.tif	<p>Soil samples were collected at depths of 0-5, 5-10, 15-20, and 25-30 cm across four Australian ecosystems, with four biological replicates for each depth (n = 64). Boxplots show median, lower and upper quartile, and minimum and maximum values.</p>
<b>Extended Data Fig. 8</b>	Comparison of the theoretical and measured population of trace gas oxidizers in soils,	Extended-Data-Figure-8.tif	<p>Shown are four biological replicates collected at depths of 0-5, 5-10, 15-20, and 25-30 cm across four Australian ecosystems (n = 64). The theoretical populations, in line with the methods of Conrad<sup>69</sup>, were estimated from thermodynamic modeling and measured <i>ex situ</i> oxidation rates assuming a single maintenance energy for all cells. Two different values were used: <math>8.9 \times 10^{-15} \text{ W cell}^{-1}</math> as assumed by Conrad<sup>69</sup> and <math>1.9 \times 10^{-15} \text{ W cell}^{-1}</math> as the median of metabolic rates reported by DeLong et al<sup>61</sup>. The measured populations of trace gas oxidizers were determined from 16S rRNA gene copy numbers and the relative abundance of the respective functional gene as determined from metagenomics. The ratio of theoretical vs. measured population is equivalent to dividing our calculated power per cell by the respective maintenance energy above. Boxplots show median, lower and upper quartile, and</p>

			minimum and maximum values.
<b>Extended Data Fig. 9</b>	Maximum-likelihood tree of amino acid sequences of ribulose 1,5-bisphosphate carboxylase / oxygenase (RbcL), a marker for carbon fixation through the Calvin-Benson cycle.	Extended-Data-Figure-9.tif	The tree shows 68 sequences from soil metagenome-assembled genomes (blue) alongside 126 representative sequences from NCBI reference genomes (black). The subtype of each reference sequence is denoted. The tree was constructed using the JTT matrix-based model, used all sites, and was bootstrapped with 50 replicates and midpoint-rooted. Source data is provided in Newick (nwk) format.
<b>Extended Data Fig. 10</b>	Summary of the processes and mediators of trace gas cycling at the soil-atmosphere interface.	Extended-Data-Figure-10.tif	The soil bacterial phyla capable of consuming the trace gases H <sub>2</sub> ( <i>via</i> group 1 and 2 [NiFe]-hydrogenases), CO ( <i>via</i> form I carbon monoxide dehydrogenases), and CH <sub>4</sub> ( <i>via</i> particulate methane monooxygenases) are shown. They are listed in order of the number of metagenome-assembled genomes recovered (black for highly abundant and grey for less abundant trace gas oxidizers). The downward arrows show the net consumption of atmospheric H <sub>2</sub> , CO, and CH <sub>4</sub> by soil bacteria. The upward arrows show processes that result in endogenous production and internal cycling of trace gases. The green boxes indicate biotic processes, whereas the brown boxes indicate abiotic processes.

6

7 **2. Supplementary Information:**

8

9 **A. Flat Files**

10

11

<b>Item</b>	<b>Present?</b>	<b>Filename</b> This should be the name	<b>A brief, numerical description of file contents.</b> i.e.: <i>Supplementary Figures 1-4, Supplementary Discussion, and</i>
-------------	-----------------	--	--

		the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	<i>Supplementary Tables 1-4.</i>
<b>Supplementary Information</b>	Yes	Supplementary information.pdf	Supplementary Tables 1 to 12 Supplementary Figures 1 to 10
<b>Reporting Summary</b>	Yes	AB Rep Summary.pdf	
<b>Peer Review Information</b>	Yes	Greening_TPR File.pdf	

12

13

14

## B. Additional Supplementary Files

15

<b>Type</b>	<b>Number</b> If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	<b>Filename</b> This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: <i>Smith_Supplementary Video 1.mov</i>	<b>Legend or Descriptive Caption</b> Describe the contents of the file
Supplementary Table		Supplementary Tables 1-12.xlsx	Supplementary Tables 1 to 12 in XLSX format.
Supplementary Data		Phylogenetic tree newick files.zip	Raw phylogenetic trees in NWK format for Figures 2a, 2b, ED Figure 3, ED Figure 5, ED Figure 9, and Supplementary Figures 1 to 10.

16

17 **3. Source Data**

18

<b>Figure</b>	<b>Filename</b> This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: <i>Smith_Source Data Fig1.xls</i> , or <i>Smith_Unmodified Gels_Fig1.pdf</i>	<b>Data description</b> i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
<b>Source Data Fig. 1</b>	Source Data Fig 1.xlsx	Numerical source data
<b>Source Data Fig. 3</b>	Source Data Fig 3.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 1</b>	Source Data Extended Data Fig 1.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 2</b>	Source Data Extended Data Fig 2.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 4</b>	Source Data Extended Data Fig 4.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 6</b>	Source Data Extended Data Fig 6.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 7</b>	Source Data Extended Data Fig 7.xlsx	Numerical source data
<b>Source Data Extended Data Fig. 8</b>	Source Data Extended Data Fig 8.xlsx	Numerical source data

19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

## Trace gas oxidizers are widespread and active members of soil microbial communities

Sean K. Bay<sup>1,2</sup>, Xiyang Dong<sup>3</sup>, James A. Bradley<sup>4,5</sup>, Pok Man Leung<sup>1,2</sup>, Rhys Grinter<sup>2</sup>, Thanavit Jirapanjawat<sup>1,2</sup>, Stefan K. Arndt<sup>6</sup>, Perran L.M. Cook<sup>7</sup>, Douglas LaRowe<sup>8</sup>, Philipp A. Nauer<sup>7</sup>, Eleonora Chiri<sup>1,2</sup>\*, Chris Greening<sup>1,2</sup>\*

<sup>1</sup>School of Biological Sciences, Monash University, Clayton, VIC 3800, Australia

<sup>2</sup>Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

<sup>3</sup>School of Marine Sciences, Sun Yat-Sen University, Zhuhai 519082, China

<sup>4</sup>School of Geography, Queen Mary University of London, Bethnal Green, London E1 4NS, United Kingdom

<sup>5</sup>Interface Geochemistry, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.

<sup>6</sup>School of Ecosystem and Forest Sciences, University of Melbourne, Richmond, VIC 3121, Australia

<sup>7</sup>School of Chemistry, Monash University, Clayton VIC 3800, Australia

<sup>8</sup>Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089, United States

### Correspondence:

\* Assoc Prof Chris Greening, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia

Email: [chris.greening@monash.edu](mailto:chris.greening@monash.edu), Ph: +61 451 085 339, ORCID: 0000-0001-7616-0594

\* Dr Eleonora Chiri, Monash University, Department of Microbiology, Innovation Walk, Clayton, VIC 3800, Australia

Email: [eleonora.chiri@monash.edu](mailto:eleonora.chiri@monash.edu), Ph: +61 3 9902 0123, ORCID: 0000-0002-6627-0762

52 **Abstract**

53           Soil microorganisms globally are thought to be sustained primarily by organic carbon  
54 sources. Certain bacteria also consume inorganic energy sources such as trace gases, but they are  
55 presumed to be rare community members, except within some oligotrophic soils. Here we  
56 combined metagenomic, biogeochemical, and modelling approaches to determine how soil  
57 microbial communities meet energy and carbon needs. Analysis of 40 metagenomes and 757  
58 derived genomes indicated over 70% of soil bacterial taxa encode enzymes to consume inorganic  
59 energy sources. Bacteria from 19 phyla encoded enzymes to use the trace gases hydrogen and  
60 carbon monoxide as supplemental electron donors for aerobic respiration. In addition, we  
61 identified a fourth phylum (Gemmatimonadota) potentially capable of aerobic methanotrophy.  
62 Consistent with the metagenomic profiling, communities within soil profiles from diverse  
63 habitats rapidly oxidized hydrogen, carbon monoxide, and to a lesser extent methane below  
64 atmospheric concentrations. Thermodynamic modelling indicated that the power generated by  
65 oxidation of these three gases is sufficient to meet the maintenance needs of the bacterial cells  
66 capable of consuming them. Diverse bacteria also encode enzymes to use trace gases as electron  
67 donors to support carbon fixation. Altogether, these findings suggest that trace gas oxidation  
68 confers a major selective advantage in soil ecosystems, where availability of preferred organic  
69 substrates limits microbial growth. The observation that inorganic energy sources may sustain  
70 most soil bacteria also has broad implications for understanding atmospheric chemistry and  
71 microbial biodiversity in a changing world.

72

## 73 **Introduction**

74 Bacteria mediate key supporting and regulatory services in soil ecosystems worldwide <sup>1,2</sup>.  
75 Culture-independent surveys have shown that soils harbor abundant and diverse bacterial  
76 communities <sup>3,4</sup>, with most cells thought to be in dormant states due to pressures such as resource  
77 limitation <sup>5,6</sup>. Most soil bacteria use organic carbon derived from vegetation and other inputs as  
78 energy and carbon sources <sup>1,7</sup>; reflecting this, isolates from all nine of the most abundant phyla in  
79 soils are aerobic organotrophs <sup>8,9</sup>. However, various bacteria have been characterized that use  
80 inorganic energy sources, including molecular hydrogen (H<sub>2</sub>), carbon monoxide (CO), sulfide,  
81 thiosulfate, ammonia, and nitrite <sup>10,11</sup>. While these lithotrophs have key ecological and  
82 biogeochemical roles <sup>1,10</sup>, they are generally thought to be minor community members compared  
83 to organotrophs <sup>1,12</sup>.

84 Emerging evidence suggests the trace gases H<sub>2</sub>, CO, and to a lesser extent methane (CH<sub>4</sub>)  
85 are particularly important energy sources for bacterial growth and survival. These three gases are  
86 ubiquitously available in the atmosphere at average global concentrations of 0.53, 0.10, and 1.85  
87 parts per million (ppmv) respectively, with local CO levels varying greatly due to natural and  
88 anthropogenic processes <sup>13-16</sup>. These compounds are also produced within soils: H<sub>2</sub> through  
89 biological nitrogen fixation and fermentation <sup>13,17,18</sup>, CO primarily through abiotic thermal and  
90 photochemical processes <sup>15,19</sup>, and CH<sub>4</sub> primarily through methanogenesis <sup>16</sup>. Biogeochemically,  
91 soil bacteria are major sinks of reduced trace gases, accounting for the net loss of approximately  
92 75% atmospheric H<sub>2</sub> (70 teragrams per year) <sup>13,17</sup>, 10% atmospheric CO (250 teragrams per year)  
93 <sup>15,20</sup>, and 5% CH<sub>4</sub> when excluding wetlands (30 teragrams per year) <sup>16,21</sup>. Microorganisms have  
94 evolved specialized metalloenzymes, namely uptake [NiFe]-hydrogenases, [MoCu]-carbon  
95 monoxide dehydrogenases, and methane monooxygenases, to oxidize these gases, including  
96 below atmospheric concentrations <sup>21-24</sup>. Some bacteria sustain autotrophic or mixotrophic growth  
97 by using electrons derived from H<sub>2</sub> and CO to drive aerobic respiration and carbon fixation  
98 through various pathways <sup>25-28</sup>. In addition, bacterial cultures from four phyla (Actinobacteriota,  
99 Proteobacteria, Acidobacteriota, Chloroflexota) have been shown to switch from growth on  
100 organic carbon to persistence on these trace gases in response to carbon starvation <sup>22-24,29,30</sup>.  
101 Genetic studies have shown that such metabolic flexibility provides a selective advantage under  
102 conditions of low resource availability or stability <sup>18,24,31-33</sup>.

103 Despite these advances, we lack a holistic understanding of the relative contributions of  
104 trace gases and other compounds in sustaining the energy and carbon needs of soil communities  
105 <sup>34</sup>. Some have hypothesized that trace gas oxidizers may be widespread community members  
106 <sup>20,24,26</sup>. However, previous soil surveys reported that atmospheric H<sub>2</sub> and CO oxidizers are



107 members of the rare biosphere, each comprising just 1% of the total bacterial community,  
108 similarly to methane-oxidizing bacteria (methanotrophs)<sup>12,35,36</sup>. Notable exceptions include some  
109 extreme environments, such as Antarctic deserts, where trace gas oxidizers are major primary  
110 producers in soils with otherwise low photosynthetic input<sup>37-39</sup>. Here, we combined  
111 metagenomic, biogeochemical, and thermodynamic modelling approaches to profile metabolic  
112 strategies of soil bacteria. We reveal facultative trace gas oxidizers are highly abundant and  
113 phylogenetically diverse members of soil microbiomes. Moreover, consumption of trace gases  
114 by these bacteria would be sufficient to sustain their basal energy needs.

## 115 **Results and Discussion**

### 116 **Most soil bacteria harbor the flexibility to oxidize organic and inorganic energy sources**

117 First, we sequenced metagenomes from depth profiles of four Australian soil ecosystems  
118 (wetland, grassland, forest, dryland) (**Supplementary Tables 1 & 2**) and analyzed metagenomes  
119 from eight further global sites (**Supplementary Table 2**). Assembly and binning yielded 757  
120 high- or medium-quality metagenome-assembled genomes (MAGs) spanning 26 bacterial and  
121 five archaeal phyla (**Supplementary Table 3**). Consistent with the community composition of  
122 the sites (**Extended Data Fig. 1; Supplementary Table 4**), half of the genomes affiliate with  
123 globally dominant soil phyla<sup>3,8</sup> Actinobacteriota, Proteobacteria, and Acidobacteriota.

124 We used comprehensive reference databases (**Supplementary Table 5**) to search key  
125 metabolic genes in metagenomic short reads and derived genomes (**Fig. 1; Supplementary**  
126 **Tables 6 & 7**). As anticipated<sup>1</sup>, most bacteria have the capability to conserve energy through  
127 aerobic respiration of organic compounds. More surprisingly, many of these bacteria also  
128 encoded the key enzymes (group 1 and 2 [NiFe]-hydrogenases and form I CO dehydrogenases)  
129 to consume trace gases H<sub>2</sub> and CO: an average of 39% and 56% of the bacterial community  
130 based on short reads, and 31% and 26% of assembled genomes. Also widespread was the  
131 capacity for sulfide (28% community / 22% genomes) and thiosulfate (6.2% / 11%) oxidation,  
132 whereas nitrite (4.6% / 3.7%), ammonium (1.5% / 2.4%), and methane (1.1% / 0.37%) oxidation  
133 and photosynthesis (1.3% / 1.5%) were more restricted traits (**Fig. 1**). Altogether, on average  
134 72% of bacterial genomes encoded enzymes to acquire energy *via* inorganic substrates. For  
135 electron acceptor utilization, most bacteria encoded terminal oxidases for aerobic respiration and  
136 many were predicted to conserve energy by nitrate reduction (25% / 14%), nitrite reduction (21%  
137 / 21%), and hydrogenogenic fermentation (29% / 21%) (**Fig. 1**). Consistent with these results, for  
138 three global sites with accompanying metatranscriptome data (**Supplementary Table 2**), read  
139 mapping confirmed the most expressed respiratory genes encoded terminal oxidases (av. 594  
140 transcripts per million, TPM) and enzymes for oxidation of NADH (198 TPM), H<sub>2</sub> (126 TPM),  
141 CO (75 TPM), and ammonium (223 TPM) (**Extended Data Fig. 2; Supplementary Table 8**).  
142 Altogether, these observations suggest that a large proportion of soil bacteria are metabolically  
143 flexible organoheterotrophs, which encode and express enzymes to use alternative energy  
144 sources such as trace gases to supplement energy needs.

145 **Diverse bacterial phyla are capable of oxidizing H<sub>2</sub>, CO, and CH<sub>4</sub> in different soil**  
146 **ecosystems**

147 To gain insight into phylogenetic diversity of trace gas oxidizers, we generated  
148 phylogenetic trees of the uptake hydrogenases and CO dehydrogenases from binned and  
149 unbinned assembled sequences (**Supplementary Table 9**). 19 phyla encoded one or both  
150 enzymes, including all nine dominant soil phyla (Acidobacteriota, Actinobacteriota,  
151 Bacteroidota, Chloroflexota, Firmicutes, Gemmatimonadota, Planctomycetota, Proteobacteria,  
152 Verrucomicrobiota)<sup>3,8</sup> and six candidate phyla (Binatota, Dormibacterota, Eremiobacterota,  
153 Methyloirabilota, Tectomicrobia, UBP7). Among the uptake hydrogenases, 49% of the 1522  
154 hits affiliated with group 1h [NiFe]-hydrogenases. This relatively recently discovered high-  
155 affinity enzyme supports persistence of organotrophic bacteria by oxidizing atmospheric H<sub>2</sub>  
156 <sup>22,23,26,30,40,41</sup>. The phylogenetic tree revealed an unprecedented sequence diversity and broad  
157 taxonomic distribution of this subgroup, with sequences affiliated with 13 phyla (**Fig. 2a;**  
158 **Supplementary Fig. 1**). Also widespread were various radiations within the group 1d (8.9%), 1f  
159 (19%), and 2a (3.6%) [NiFe]-hydrogenases known to support aerobic hydrogen oxidation  
160 <sup>23,26,28,42,43</sup>, as well as a subclade of the group 1c hydrogenases (13%) encoded by  
161 Acidobacteriota, Planctomycetota, and Gemmatimonadota (**Fig. 2a; Supplementary Figs. 2 &**  
162 **3**). For CO dehydrogenases, we restricted our searches to the canonical form I enzyme using  
163 recently described reference databases<sup>24</sup>, given there is no strong evidence that CO is the  
164 physiological substrate of the related form II enzyme<sup>20,24,29,44</sup>. The form I CO dehydrogenase  
165 was similarly prevalent across the MAGs; the tree shows large actinobacterial and  
166 proteobacterial clades flanking a central mixed clade containing sequences from 11 different  
167 phyla (**Fig. 2b; Supplementary Fig. 4**). In line with previous inferences<sup>24,26</sup>, both trees suggest  
168 enzymes for trace gas oxidation were horizontally disseminated between bacteria on multiple  
169 occasions.

170 Phylogenetic analysis also validated the presence and affiliations of sulfide, thiosulfate,  
171 ammonia, nitrite, and methane oxidizers (**Supplementary Figs. 5 to 10; Extended Data Fig. 3**).  
172 These included genomes of soil comammox bacteria<sup>45,46</sup> and a potential nitrite-oxidizing  
173 bacterium within recently reported candidate phylum Eremiobacterota<sup>37</sup>. A major finding was  
174 the recovery of a high-quality Gemmatimonadota bin predicted to use methane, hydrogen, and  
175 acetate as energy sources. Its genome encodes a putative particulate methane monooxygenase  
176 that forms a sister lineage to those of the recently discovered *Candidatus* Methyloirabilota  
177 methanotrophs (sharing ~60% sequence identity) and corresponds to amplicon sequences  
178 previously identified for the ubiquitous uncultivated putative methanotroph lineage TUSC

179 (Tropical Upland Soil Cluster)<sup>47,48</sup>. Protein modelling suggests that this enzyme shares a  
180 common tertiary structure and copper-binding residues with canonical particulate methane  
181 monooxygenases (**Extended Data Fig. 3**). Altogether, these results suggest that  
182 Gemmatimonadota is a fourth phylum<sup>49–51</sup> capable of aerobic methanotrophy, though cultivation  
183 and biogeochemical studies are required to confirm its activity. We propose this bacterium is  
184 named *Candidatus Methylootropicum kingii* (**Etymological Information**).

185

## 186 **Trace gas oxidizers are highly active across different soil habitats and depths**

187 Reflecting their abundance and diversity, we confirmed that trace gas oxidizers are active  
188 in soil communities of all investigated Australian soils. *In situ* concentrations and soil-  
189 atmosphere fluxes of H<sub>2</sub>, CO, and CH<sub>4</sub> were measured across the four sites in biological  
190 quadruplicate using a sensitive gas chromatograph (**Supplementary Table 10**). All three gases  
191 were present within typical mixing ratios<sup>13,15,16</sup> at the soil-atmosphere interface and top 20 cm of  
192 each soil (**Fig. 3a**), confirming these substrates are available to bacteria. Gas concentrations  
193 decreased with depth at some sites (**Extended Data Fig. 4**), potentially indicating microbial  
194 consumption. There was also potential evidence of endogenous gas production within some soil  
195 profiles: the elevated H<sub>2</sub> concentrations in deeper wetland soils may result from activity of the  
196 numerous hydrogenogenic fermenters (33% community / 32% genomes; **Extended Data Fig. 5**)  
197 under hypoxic conditions<sup>18</sup>, whereas the elevated gas levels in the dryland soils are consistent  
198 with photochemical and thermal production processes<sup>52</sup>. In turn, endogenous production would  
199 result in some cryptic gas cycling within the soil profile, though this remains to be  
200 experimentally verified. In line with the abundance of H<sub>2</sub> oxidizers, net *in situ* H<sub>2</sub> uptake was  
201 observed by soil communities at all sites and consumption was particularly rapid at the  
202 grassland, forest, and wetland sites (**Fig. 3b**); ambient H<sub>2</sub> in the flux chambers was oxidized by  
203 underlying soil down to a threshold of ~0.1 ppmv within a few minutes, resulting in typical  
204 fluxes of -10 to -40 nmol m<sup>-2</sup> s<sup>-1</sup>. In contrast, no significant fluxes of CO were observed; the high  
205 rates of photochemical CO production known to occur during daylight potentially obscured  
206 underlying microbial consumption<sup>19,53</sup>, though this remains to be experimentally confirmed.  
207 Reflecting global patterns<sup>54</sup>, with the exception of forest sites, CH<sub>4</sub> uptake fluxes were low in  
208 comparison with H<sub>2</sub>. Control experiments confirmed the flux chambers were gas-tight and did  
209 not mediate significant gas production or consumption during the timecourse of the flux  
210 measurements (**Extended Data Fig. 4**). Overall, the trace gas fluxes observed are much lower  
211 than typically reported for heterotrophic respiration<sup>7</sup>, although endogenous soil production  
212 potentially obscures much consumption.

213 We subsequently measured *ex situ* oxidation rates by incubating samples from soil depth  
214 profiles in serum vials containing ambient air headspaces (**Supplementary Table 10**). Under  
215 these conditions, biological gas consumption could be accurately quantified given no significant  
216 abiotic production of these gases was detected (**Extended Data Fig. 6**). Atmospheric H<sub>2</sub> and CO  
217 were both rapidly oxidized by forest, grassland, and wetland soils at rates within fivefold of  
218 previous literature<sup>10,55-58</sup>. Bulk rates were significantly higher for topsoils (0-5 and 5-10 cm)  
219 compared to deeper soils (15-20 and 25-30 cm) (**Supplementary Table 10; Extended Data Fig.**  
220 **6**), though cell-specific rates (normalized to estimated number of respective trace gas oxidizing  
221 cells; **Extended Data Fig. 7**) showed few significant variations with depth (**Fig. 3c;**  
222 **Supplementary Table 11**). Across these sites, atmospheric CH<sub>4</sub> oxidation occurred at 76-fold  
223 and 35-fold lower rates than H<sub>2</sub> and CO on average, with the most rapid consumption again  
224 occurring at the forest site. Such observations are consistent with the much higher levels of  
225 uptake hydrogenases and CO dehydrogenases compared to methane monooxygenases in the  
226 metagenomes and derived genomes (**Fig. 1**). In line with the *in situ* observations, trace gas  
227 oxidation was slowest for dryland soils. This possibly reflects that, as observed in previous  
228 laboratory studies, low water content inhibits trace gas uptake<sup>44,45</sup>. Contrary to the pattern  
229 observed in other habitats, deeper dryland soils consumed trace gases more rapidly, perhaps  
230 reflecting their higher measured water content (**Supplementary Table 1**). Experimentally  
231 wetting dryland soils, in order to simulate rainfall events, enhanced rates by an average of 60-  
232 fold (**Fig. 3c; Extended Data Fig. 6**). Together, these activity-based measurements complement  
233 the metagenomic analysis by confirming soil communities contain highly active H<sub>2</sub>, CO, and  
234 sometimes CH<sub>4</sub> oxidizers. While net *in situ* gas fluxes were often low or variable, especially for  
235 CO, the rapid *ex situ* rates and previous reports of endogenous gas production<sup>18,19,52,53,59</sup> suggest  
236 high levels of internal gas cycling likely occurs.

237

### 238 **Trace gases enable persistence of most soil bacteria and may support mixotrophic growth**

239 Trace gas oxidation is likely to primarily support the persistence of organotrophic  
240 bacteria. Most sequenced soil bacteria encoded uptake hydrogenases and CO dehydrogenases  
241 together with the genes for organotrophy. Based on observations from pure culture studies,  
242 bacteria consume trace gases to conserve energy for cellular maintenance when organic carbon  
243 supplies are limiting for growth<sup>3-5,30</sup>. We used thermodynamic modelling to predict the amount  
244 of power (i.e. energy per time in J s<sup>-1</sup> (W)) per cell (**Supplementary Table 12**) that could be  
245 generated based on the *ex situ* oxidation rates (**Extended Data Fig. 6**) and the total number of  
246 putative trace gas oxidizers detected per gram of dry soil (**Fig. 1 & Extended Data Fig. 7**).

247 Oxidation rates were sufficient to generate an average power of  $2.2 \times 10^{-15}$  W per H<sub>2</sub>-oxidizing  
248 cell,  $2.0 \times 10^{-15}$  W per CO-oxidizing cell, and  $1.1 \times 10^{-13}$  W per CH<sub>4</sub>-oxidizing cell, and trends  
249 were similar across habitats and depths with the exception of the dryland soils (**Fig. 3d &**  
250 **Supplementary Table 12**). The calculated cell-specific power values are within the broad range  
251 of the maintenance and endogenous energy requirements previously reported for various  
252 laboratory-grown cultures of aerobic heterotrophs (average  $1.9 \times 10^{-15}$  W per cell<sup>60</sup>; range  $10^{-12}$   
253 to  $10^{-17}$  W per cell<sup>60-63</sup>) and greatly exceed the *in situ* maintenance energies of aerobic  
254 heterotrophs in highly energy-limited ecosystems ( $10^{-17}$  to  $10^{-19}$  W per cell<sup>64-66</sup>). Previous  
255 studies estimated the theoretical maximum population of trace gas oxidizers based on observed  
256 oxidation rates and the assumption of a single static maintenance energy<sup>67-69</sup>. Although our  
257 experimentally observed population is in reasonable agreement with such estimates, results are  
258 sensitive to the assumed maintenance energy (**Extended Data Fig. 8**). Literature data and our  
259 cell-specific power calculations based on empirical parameters suggest that metabolic rates are  
260 highly variable between microbial populations, metabolic strategies, and soil habitat. Therefore,  
261 the use of a single maintenance energy to estimate the fraction of the population that can be  
262 supported by trace gas oxidation may be too simplistic. It should also be noted that our estimates  
263 did not account for relic DNA<sup>70</sup>, inactive cells, and cryptic gas cycling. Thus, gross oxidation  
264 rates and available power per active cell may be considerably higher *in situ*. Altogether, this  
265 suggests that trace gases alone can theoretically sustain the persistence of the entire gas-  
266 consuming community.

267 Trace gases may also support mixotrophic growth of a significant proportion of bacteria  
268 across these soil habitats. Our analysis indicates that most community members, including trace  
269 gas oxidizers, acquire carbon heterotrophically. However, some bacteria (13% community / 12%  
270 genomes) encoded the capacity to assimilate CO<sub>2</sub> through the Calvin-Benson cycle (**Fig. 1**).  
271 Genome-resolved analysis indicated most of these autotrophs were capable of oxidizing H<sub>2</sub>  
272 (79%) and/or CO (63%), with some predicted to mediate sulfide oxidation (32%), thiosulfate  
273 oxidation (17%), or photosynthesis (11%) (**Supplementary Table 7**). Putative autotrophic H<sub>2</sub>  
274 and CO oxidizers were patchily distributed among seven dominant soil phyla and candidate  
275 phylum Dormibacterota, suggesting autotrophy has been acquired multiple times (**Fig. 1**).  
276 Consistently, phylogenetic analysis showed that the 93% of the binned ribulose 1,5-bisphosphate  
277 carboxylase / oxygenase (RuBisCO / RbcL) hits were from clades known to support  
278 chemosynthesis rather than photosynthesis (**Extended Data Fig. 9**). Reads for other CO<sub>2</sub> fixation  
279 pathways were also detected, but with exception of the peatland sample, were encoded by less  
280 than 1% of the total community (**Fig. 1; Supplementary Table 6**). Individual atmospheric trace  
281 gases are thought to be insufficiently concentrated to sustain growth as the sole energy source

282 <sup>10,67</sup>, though there is increasing evidence that they can sustain mixotrophic growth <sup>27,28,71</sup>.  
283 Moreover, autotrophic growth may be favorable in environments where there are significant  
284 rates of edaphic gas production, for example due to hydrogenogenic fermentation (**Extended**  
285 **Data Fig. 5**) during hypoxia <sup>28</sup>. Remarkably, this suggests that the process of atmospheric  
286 chemosynthesis recently discovered in Antarctic deserts <sup>9</sup> extends to other habitats.

287

## 288 **Conclusions and Outlook**

289 Overall, bacteria with the metabolic flexibility to use both organic and inorganic energy  
290 sources are likely to have a selective advantage in soil environments. In most soils, organic  
291 carbon is the main factor limiting microbial growth <sup>13,49</sup>; this reflects the inherent spatiotemporal  
292 variability in organic carbon availability of soils, together with the recalcitrance of many organic  
293 polymers and the intense competition for more degradable compounds <sup>1</sup>. Thus, the ability to  
294 consume alternative energy sources is likely to be critical for adaptation and resilience of many  
295 taxa. H<sub>2</sub> and CO are ideal compounds in this regard given they are readily available from both  
296 atmospheric and edaphic sources; likewise, their high energy content, low activation energy, and  
297 diffusibility into microbial cells make these gases dependable for persistence <sup>31,50</sup>.

298 Our findings in turn suggest that oxidation of these trace gases is a generalist process,  
299 rather than a specialist one as previously suggested <sup>6</sup>. We provide multiple lines of evidence that  
300 H<sub>2</sub> and CO oxidizers are abundant, diverse, and active across different soil habitats and depths.  
301 These findings strikingly contrast with some previous reports that trace gas oxidizers comprise  
302 just 1% of the community <sup>12,35</sup>. Such discrepancies might reflect that previous work on  
303 atmospheric H<sub>2</sub> oxidizers relied on non-degenerate quantitative PCR primers that only capture a  
304 small proportion of the total diversity of soil uptake hydrogenases; indeed, we observed high  
305 proportions of high-affinity hydrogenases (30% community / 40% genomes) in the Canadian  
306 cropland metagenome where H<sub>2</sub> oxidizers were previously inferred to be minor community  
307 members <sup>6,7</sup>. Moreover, while previous genome surveys indicate Actinobacteriota predominantly  
308 mediate atmospheric H<sub>2</sub> oxidation <sup>33,37</sup>, our analysis indicates all major bacterial soil phyla can  
309 mediate this process. However, an important caveat is that our findings are based primarily on  
310 genomic inferences and bulk rates, meaning it is unclear if all putative trace gas oxidizers were  
311 active during the experiments. Such metabolically flexible generalists co-exist with more  
312 specialist taxa that are predicted to use more niche substrates, including methane, ammonia, and  
313 nitrite. The finding that some Gemmatimonadota encode genes for aerobic methane oxidation  
314 further indicates the diversity of trace gas oxidizers in soils. Our results also reveal an

315 unexpected diversity and abundance of chemoautotrophs in soils, while hinting at roles for  
316 sulfur-based compounds as further supplementary energy sources.

317 The thermodynamic calculations suggest that trace gas oxidation is sufficient to sustain  
318 the persistence of all trace gas oxidizers and thus most bacteria in soils. Power per cell values are  
319 well within the range of maintenance or endogenous energies typically reported for pure cultures,  
320 and exceed the range for energy-limited ecosystems<sup>60-64</sup>. Nevertheless, there is likely to be  
321 considerable heterogeneity in the rates that individual cells oxidize trace gases and the way they  
322 allocate energy derived from this process. For example, it is plausible that whereas some cells  
323 encoding uptake hydrogenases minimally use H<sub>2</sub>, many others will depend on this gas for  
324 survival and yet others may use it to support growth. It is also critical to note that most bacteria  
325 harboring genes for trace gas oxidation are predicted to be metabolically flexible. Many of them  
326 are likely to simultaneously consume organic and inorganic energy sources, or indeed several  
327 trace gases together as recently inferred for a methanotroph predicted to grow on atmospheric  
328 energy sources alone<sup>27</sup>, in order to grow or persist mixotrophically. Innovative studies are  
329 required to gain a detailed perspective on how trace gases are differentially used within and  
330 between bacterial populations and cells.

331 More broadly, the extensive soil-atmosphere interaction described here appears to be a  
332 key regulator of soil biodiversity and atmospheric chemistry. As summarized in **Extended Data**  
333 **Fig. 10**, soil bacteria consume trace gases of atmospheric, biological, and geochemical origin.  
334 Given trace gases are major energy sources sustaining the dormant soil majority, their oxidation  
335 will in turn influence wider community ecology; indeed, it is well-established that dormancy  
336 contributes to the maintenance of microbial biodiversity and the resilience of soil communities  
337<sup>17</sup>. Moreover, the observation that trace gas oxidizers are more numerous and diverse than  
338 previously thought suggests the global atmospheric sinks of H<sub>2</sub> and CO are relatively resilient.  
339 This may explain why, despite high anthropogenic emissions of H<sub>2</sub>, global mixing ratios have  
340 remained quite stable<sup>26</sup>. Nevertheless, various human activities could undermine this soil-  
341 atmosphere interaction, for example through changing soil properties (e.g. through agricultural  
342 practices and desertification) and atmospheric composition (e.g. through urban pollution or a  
343 hydrogen economy<sup>51</sup>), with potential ecological and biogeochemical ramifications. Further  
344 studies are required to understand what physicochemical factors influence the abundance and  
345 activities of trace gas oxidizers, and how they respond to local and global change. More work is  
346 also needed to understand the role of chemosynthetic CO<sub>2</sub> fixation by soil trace gas oxidizers,  
347 which appear to include the most abundant bacterial autotrophs in aerated soils, and in turn  
348 determine whether this process serves as a significant CO<sub>2</sub> sink.





## 350 **Methods**

351

352 **Site description and sampling.** Sampling was conducted in four sites within Australia: (i)  
353 wetland (-37.908647°, 145.139591°; Jock Marshall Reserve, Clayton, VIC; JMR), (ii) forest (-  
354 37.446379°, 144.470078°; Wombat State Forest, VIC; WSF), (iii) grassland (-37.926514°,  
355 145.313136°; Don Bosco Grassland, Lysterfield, VIC; DBG), and (iv) dryland (-23.977765°,  
356 133.712524°; 25.5 km south of Alice Springs, NT; ASD). The sites were sampled on December  
357 19 2018, January 5 2019, January 21 2019, and January 29 2019 respectively. At each site (~100  
358 m<sup>2</sup>), four sampling plots (~1 m<sup>2</sup>) were selected. At each plot, *in situ* depth-resolved gas  
359 concentrations profiles and soil-atmosphere fluxes were measured. In addition, a core was used to  
360 collect soil samples at four depths (0-5, 5-10, 15-20, and 25-30 cm) for *ex situ* oxidation  
361 measurements, physicochemical analysis, and DNA extractions to perform quantitative  
362 polymerase chain reactions (qPCR) and metagenomic sequencing. All sampling occurred during  
363 daylight hours and dry weather conditions, gas samples were processed within 48 h of collection,  
364 and soil samples were incubated for *ex situ* oxidation measurements within 24 h of collection.

365

366 **Sampling and measurement of soil gas profiles.** Depth-resolved *in situ* gas concentrations of  
367 H<sub>2</sub>, CO, and CH<sub>4</sub> were measured using stainless steel capillaries of 1 mm internal diameter fitted  
368 with a Luer Lock and Discifix three-way stopcock. Prior to sampling, each capillary was inserted  
369 into the soil at depth intervals of 2, 4, 6, 8, 10, and 16 cm. All capillaries were installed  
370 simultaneously, arranged in a hexagonal grid (diameter ~20 cm), and left to equilibrate for ~30  
371 min. Samples of 15 mL were withdrawn using a 20 mL gas-tight Terumo syringe. A gas sample  
372 was also collected at the soil-atmosphere interface (0 cm). All gas samples were stored in  
373 evacuated 12 mL glass exetainers sealed with rubber septum lids and analyzed using gas  
374 chromatography. Gas concentrations in samples were measured by gas chromatography using a  
375 pulsed discharge helium ionization detector (model TGA-6791-W-4U-2, Valco Instruments  
376 Company Inc.) as previously described <sup>72</sup>. Samples were calibrated against H<sub>2</sub>, CO and CH<sub>4</sub>  
377 standards that were prepared using ultra-pure concentrations of each gas (1% in N<sub>2</sub> gas cylinder,  
378 99.999% pure, Air Liquide Australia) down to the limit of quantification (H<sub>2</sub> 20 ppbv, CO 90  
379 ppbv, CH<sub>4</sub> 500 ppbv. Pressurized air (Air Liquide Australia) with known trace gas concentrations  
380 was used as an internal reference standard.

381

382 **Measurement of soil-atmosphere gas fluxes.** *In situ* soil-atmosphere fluxes of H<sub>2</sub>, CO, and CH<sub>4</sub>  
383 were measured using static flux chambers. The chamber consisted of a polyvinylchloride (PVC)  
384 pipe of 20 cm height and 15 cm diameter with a threaded access cap. The cap was fitted with a

385 gastight O-ring, two butyl rubber septa (one for air sampling and one for a thermometer), and an  
386 axial fan on the inside to promote internal mixing. At each plot, a PVC base collar of 10 cm  
387 height and 14.8 cm diameter was inserted ~5 cm into the soil and left to equilibrate for ~30  
388 minutes prior to sampling to reduce lateral gas fluxes. Once the chamber was fitted over the  
389 collar, the cap was closed and the axial fan was started. Three consecutive gas measurements  
390 were taken at approximately one-minute intervals, followed by either four or five measurements  
391 at approximately five-minute intervals. For each measurement, 15 mL of gas was collected using  
392 a gas-tight 20 mL Terumo syringe fitted with a Luer Lock and Discifix three-way stopcock and  
393 measured by gas chromatography as described above. Control gas measurements of ambient air  
394 were taken directly before, during, and after sampling. The temperature of the chamber, ambient  
395 air, and soil were monitored throughout. Concentrations were then converted to  $\text{nmol m}^{-3}$  at  
396 ambient pressure and temperature using the ideal gas law. Atmospheric flux ( $J_{\text{atm}}$ ) was calculated  
397 from the concentration ( $C_0$ ) and rate change ( $dC/dt$ ) at chamber deployment using a linear  
398 regression (eq. 1) and an exponential model fit (eq. 2 and eq. 3) <sup>59</sup> for each chamber  
399 measurement:

400

$$401 \quad J_{\text{atm}} = h \frac{dC}{dt} \quad (\text{EQ. 1})$$

402

$$403 \quad C_{(t)} = \frac{P}{k} + \left( C_0 - \frac{P}{k} \right) e^{(-kt)} \quad (\text{EQ. 2})$$

404

$$405 \quad J_{\text{atm}} = h(P - kC_0) \quad (\text{EQ. 3})$$

406

407 For the exponential model, parameters  $P$  and  $k$  were fitted via non-linear regression using the  
408 `nls()` function in R.  $h$  denotes the effective chamber height in m. The best model was chosen  
409 according to the lower Akaike information criterion (AIC). Conservative flux detection limits  
410 incorporating errors of sample handling and storage were calculated using mean and standard  
411 deviation of air samples <sup>73</sup>. Three independent incubations were conducted to control for trace  
412 gases potentially leaked from the chamber or released in the chamber headspace by plastic  
413 components due to abiotic processes. Chambers were placed on an inert stainless steel surface  
414 and sealed on the surface with silicone grease. Headspace samples from eight time points and  
415 three air samples were collected following measurement setup, incubation time, and sampling  
416 procedures described above for the *in situ* sampling of actual samples. No variation in trace gas  
417 concentration was observed during the incubation time (**Extended Data Fig. 8**).

418

419 **Soil sampling and physicochemical analysis.** Four soil cores of 30 cm depth were collected  
420 from each site (16 total). Cores were carefully extracted and immediately segmented at 5 cm  
421 depth intervals before being transferred into 50 mL Falcon tubes. At four different depth intervals  
422 (0-5, 5-10, 15-20, 25-30 cm), a subset of 10 g of soil was frozen at -20°C for community DNA  
423 extraction and a subset of 5 g was used within 24 hours for gas chromatography studies.  
424 Additional surface soil was collected using a 10 × 5 cm bulk density ring to estimate bulk density  
425 and soil water content, which was measured gravimetrically using a drying oven at 140°C. For  
426 soil chemistry analysis, surface soil samples (0-5 cm) from each plot were pooled to form one  
427 representative composite sample per site and sent to the Environmental Analysis Laboratory  
428 (EAL), Southern Cross University. In total, 21 separate soil chemical parameters were selected  
429 for analysis, based on commonly reported drivers of soil microbial composition globally. These  
430 included: soil acidity (pH), electrical conductivity (EC), effective cation exchange capacity  
431 (ECEC), total organic carbon, total nitrogen, sodium (Na), sulfur (S), phosphate (P), potassium  
432 (K), nitrate (NO<sub>3</sub><sup>-</sup>), and ammonium (NH<sub>4</sub><sup>+</sup>), as well as bioavailable manganese (Mn), copper  
433 (Cu), zinc (Zn), boron (B), aluminium (Al), iron (Fe), and silicon (Si). In addition, particle size  
434 analysis by hydrometry was performed to estimate the percentage of gravel, sand, silt, loam, and  
435 clay (**Supplementary Table 1**).

436

437 ***Ex situ* oxidation rates.** To determine the capacity of the soils to oxidise trace gases at  
438 atmospheric concentrations, 5 g soil from the four depth intervals (0-5, 5-10, 15-20, 25-30 cm)  
439 were placed in 120 mL serum vials. The headspace was repeatedly flushed with air from a  
440 pressurized cylinder (Air Liquide, Australia) to achieve headspace mixing ratios reflecting  
441 atmospheric levels (~0.5 ppmv H<sub>2</sub>, ~0.6 ppmv CO, ~1.8 ppmv CH<sub>4</sub>). Sampling commenced  
442 immediately after sealing the vial and headspace samples of 2 mL were taken at 10 min intervals  
443 for 40 minutes. While this timecourse sufficiently captured oxidation rates for most samples,  
444 additional gas samples were collected every 24 hrs for up to nine days to capture oxidation rates  
445 of soils that mediated slow gas consumption. Gas concentrations were measured by gas  
446 chromatography as described above. Heat killed soils (two 30-minute autoclave cycles at 121°C)  
447 and blank measurements (empty serum vials) were used as controls, confirming trace gas  
448 oxidation from samples occurred due to biotic processes. Given the low capacity of dryland soils  
449 to oxidize atmospheric trace gases, we simulated soil wetting conditions due to a rainfall event to  
450 determine whether soil hydration enhances activity. To do this, we used a custom Perspex collar  
451 fitted with a water-draining stainless steel woven mesh (0.17 mm) and a water-catching tray.  
452 Collars were sterilised using ethanol and soils (5 g) were placed in the centre of the mesh surface.  
453 Soils were watered until fully saturated by repeated addition of MilliQ water. Once fully

454 saturated, each collar was sealed at the top using clingfilm to avoid evaporation and left to drain  
455 for 24 hrs in the dark. Once drained, soil samples were transferred using sterile techniques into a  
456 120 mL serum vial and gas oxidation was measured as described in the previous section. Similar  
457 to *in situ* fluxes above, reaction rates were calculated from the first five data points using first-  
458 and zero-order reaction models, with the best rate chosen according to AIC. The data was tested  
459 for normality using a Shapiro-Wilk test. To test for significant difference in oxidation rates  
460 between ecosystems and depths, a non-parametric one-way Kruskal-Wallis test was used. This  
461 was followed by a pairwise Wilcoxon Rank Sum test to test significant relationships between pairs.  
462 *p* values were adjusted using a Benjamini Hochberg correction.

463

464 **Community DNA extraction.** At each soil depth sampled per site, soils from the four plots were  
465 pooled together. For each of the 16 resultant samples, total community DNA was extracted using  
466 0.25 g soil. Extractions were performed using the MoBio PowerSoil Isolation kit according to the  
467 manufacturer's instructions. Samples were eluted in DNase- and RNase-free UltraPure Water  
468 (ThermoFisher). A sample-free negative control was also run. Nucleic acid purity and yield were  
469 measured using a NanoDrop ND-1000 spectrophotometer and a Qubit Fluorometer 2.0.

470

471 **Quantitative PCR.** Quantitative polymerase chain reactions (qPCR) were used to estimate total  
472 bacterial and archaeal biomass. The 16S rRNA gene was amplified using the degenerate primer  
473 pairs 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GGACTACNVTGGGTWTCTAAT)<sup>74</sup>.  
474 A synthetic *E. coli* 16S rRNA gene sequence in a pUC-like cloning vector (pMA plasmid;  
475 GeneArt, ThermoFisher Scientific) was used as a standard. PCR reactions were set up in each  
476 well of a 96-well plate using a LightCycler 480 SYBR Green I Master Mix. Each sample was run  
477 in triplicate and standards in duplicate on a LightCycler 480 Instrument II (Roche). The qPCR  
478 conditions were as follows: pre-incubation at 95°C for 3 min and 45 cycles of denaturation 95°C  
479 for 30 s, annealing at 54°C for 30 s, and extension at 72°C for 24 s. 16S rRNA gene copy  
480 numbers were calculated based on a standard curve constructed by plotting average Cp values of  
481 a serial dilution of the plasmid-borne standard against their copy numbers.

482

483 **Sequencing, assembly, and binning of Australian metagenomes.** Metagenomic shotgun  
484 libraries were prepared for the 16 samples using the Nextera XT DNA Sample Preparation Kit  
485 (Illumina Inc., San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq500  
486 platform with a 2 × 150 bp High Output run. Raw reads derived from the 16 metagenome  
487 libraries were quality-controlled by clipping off primers and adapters then filtering out artifacts  
488 and low-quality reads using Read\_QC module within the metaWRAP pipeline<sup>75</sup>. For each

489 ecosystem, the four quality-controlled metagenomes were co-assembled using MEGAHIT v1.1.3  
490 <sup>76</sup> (default parameters) and individually assembled using SPAdes v3.13.0 <sup>77</sup> (metaSPAdes mode,  
491 default parameters), producing five assemblies for each ecosystem. Short contigs (<1000 bp)  
492 were removed. Each assembly was binned using the binning module within the metaWRAP <sup>75</sup>  
493 pipeline (MetaBAT <sup>78</sup>, MetaBAT2 <sup>79</sup>, and MaxBin2 <sup>80</sup>). For each assembly, the three bin sets  
494 were then consolidated into a final bin set with the bin\_refinement module of metaWRAP <sup>75</sup>. For  
495 each ecosystem, the final bin sets were aggregated and de-replicated using dRep <sup>81</sup> (-comp 50 -  
496 con 10 options). Completeness, contamination, and heterogeneity of each bin were estimated  
497 using CheckM <sup>82</sup>, with medium- and high-quality bins (completeness >50%, contamination  
498 <10%) <sup>83</sup> retained for further analysis. After dereplication, a total of 93 metagenome-assembled  
499 genomes (MAGs) were obtained for the four ecosystems. Each bin was taxonomically assigned  
500 according to the Genome Taxonomy Database (GTDB; release 04-RS89) <sup>84</sup> using GTDB-tk <sup>85</sup>.

501

502 **Assembly and binning of global public soil metagenomes.** A total of 24 previously sequenced  
503 metagenomes from eight different soil ecosystems (three metagenomes for each ecosystem)  
504 (**Supplementary Table 2**) were downloaded from the Integrated Microbial Genomes database <sup>86</sup>  
505 and the NCBI Sequence Read Archive (SRA) <sup>87</sup>. These comprised: Barrow Environmental  
506 Observatory site, Barrow, Alaska, USA (Arctic Tundra, BEO) <sup>88</sup>; St. Claude, Quebec, Canada  
507 (Agricultural Land – Crop Rotation, SCQ) <sup>89</sup>; Kellogg Biological Station, Michigan, USA  
508 (Agricultural Land – Switchgrass, KBS) <sup>90</sup>; Algoma, Ontario, Canada (Coniferous Forest, ALO)  
509 <sup>91</sup>; Anza Borrego Desert, California, USA (Hot Desert, ABD) <sup>92</sup>; Department of Meta, Colombia  
510 (Tropical Peatland, DMC) <sup>93</sup>; National Park of Serra do Cipo, Brazil (Rupestrian Grassland, NPS)  
511 <sup>94</sup>; and Luquillo Experimental Forest, Rio Grande, Puerto Rico (Tropical Rainforest, LEF) <sup>95</sup>.  
512 Raw reads derived from the 24 metagenome libraries were quality-controlled by clipping off  
513 primers and adapters and filtering out artefacts and low-quality reads using Read\_QC module in  
514 the metaWRAP pipeline <sup>75</sup>. For each ecosystem, the three quality-controlled metagenomes were  
515 both co-assembled and individually assembled using MEGAHIT v1.1.3 <sup>76</sup>, producing four  
516 assemblies for each ecosystem. For the assembly process, all of them used default parameters  
517 except co-assembly of metagenomes for Kellogg Biological Station (--k-min 27). Short contigs  
518 (<1000 bp) were removed. Each assembly was binned using the binning module within the  
519 metaWRAP <sup>75</sup> options (MetaBAT <sup>78</sup>, MetaBAT2 <sup>79</sup> and MaxBin2 <sup>80</sup>) except assemblies derived  
520 from Kellogg Biological Station and Luquillo Experimental Forest where only MetaBAT2 <sup>79</sup> was  
521 used. Dereplication and quality-control of produced bins were performed as above. After  
522 dereplication, a total of 664 high- or medium-quality MAGs were obtained for the eight  
523 ecosystems. Each bin was taxonomically assigned as above.

524

525 **Functional annotation of binned and unbinned contigs.** The sequences of 43 marker genes  
526 representing energy conservation and carbon acquisition processes were retrieved from binned  
527 and unbinned contigs. Open reading frames (ORFs) were first predicted using Prodigal v.2.6.3<sup>96</sup>  
528 and genes were annotated using a combination of homology-based searches and hidden Markov  
529 model (HMM) searches. For homology-based searches, predicted ORFs were searched using  
530 DIAMOND *blastp*<sup>97</sup> against 32 custom protein databases described below. These encompassed  
531 the genes encoding ATP-citrate lyase (AclB), acetyl-CoA synthase (AcsB), ammonia  
532 monooxygenase (AmoA), anaerobic sulfite reductase (AsrA), anaerobic carbon monoxide  
533 dehydrogenase (CooS), form I aerobic carbon monoxide dehydrogenase (CoxL), dissimilatory  
534 sulfite reductase (DsrA), flavocytochrome *c* sulfide dehydrogenase (FCC), 4-hydroxybutyryl-  
535 CoA synthase of Crenarchaeota (HbsC), 4-hydroxybutyryl-CoA synthase of Thaumarchaeota  
536 (HbsT), hydrazine synthase (HzsA), malonyl-CoA reductase of Chloroflexota (Mcr),  
537 methyl/alkyl-CoM reductase (McrA), soluble methane monooxygenase (MmoX), periplasmic  
538 nitrate reductase (NapA), dissimilatory nitrate reductase (NarG), nitrogenase (NifH), copper-  
539 containing nitrite reductase (NirK), cytochrome *cd*<sub>1</sub> nitrite reductase (NirS), nitrous oxide  
540 reductase (NosZ), ammonia-forming nitrite reductase (NrfA), nitrite oxidoreductase (NxrA),  
541 particulate methane monooxygenase (PmoA), ribulose 1,5-bisphosphate carboxylase/oxygenase  
542 (RbcL), succinate dehydrogenase / fumarate reductase (SdhA / FrdA), sulfur  
543 oxygenase/reductase (Sor), thiosulfohydrolase (SoxB), sulfide-quinone oxidoreductase (Sqr), and  
544 three hydrogenase classes (NiFe-hydrogenase large subunit, FeFe-hydrogenase catalytic domain,  
545 Fe-hydrogenase). DIAMOND mapping was performed with a query coverage threshold of 80%  
546 for all databases, and a percentage identity threshold of 60% (AmoA, PmoA, MmoX, CoxL,  
547 HbsT, NxrA, RbcL) or 50% (all other databases) and e-value thresholds of 10<sup>-20</sup>. HMM searches  
548 were performed against Pfam and Tigrfam databases using the *annotate* function of EnrichM  
549 v.0.5.0 (<https://github.com/geronimp/enrichM>), with domain noise-cut-off scores as previously  
550 described<sup>98</sup>. Eleven genes encoding subunits of ATP synthase (AtpA), two NADH  
551 dehydrogenases (NuoF, NqrF), four terminal oxidases (CcoN, CoxA, CydA, CyoA), two  
552 photosystems (PsaA, PsbA), formate dehydrogenase (FdhA), and reductive dehalogenase  
553 (RdhA) were searched.

554

555 **Metabolic annotation of metagenomic short reads.** For the functional annotation of short  
556 reads, paired-end reads in each sample were stripped of adapter and barcode sequences, then  
557 contaminating PhiX and low quality sequences were removed (minimum quality score 20) using  
558 the BBDuk function of BBTools v. 36.92 (<https://sourceforge.net/projects/bbmap/>). Resultant

559 quality-filtered forward reads with lengths of at least 100 bp were searched for the presence of the  
560 43 metabolic marker genes described above using DIAMOND *blastx* algorithm <sup>99</sup>. Specifically,  
561 reads were searched against the 32 custom-made reference databases and hits from the 11 HMM  
562 searches, using a query coverage of 80% and an identity threshold of either 60% (AmoA, PmoA,  
563 MmoX, CoxL, HbsT, NxrA, RbcL) or 50% (all other databases) and a maximum e-value  
564 threshold of 10<sup>-10</sup>. Read counts were normalized to reads per kilobase per million (RPKM) and  
565 further normalized against a mean RPKM value estimated from 14 single copy ribosomal marker  
566 genes to infer the percentage of the community encoding the gene.

567

568 **Metabolic annotation of metatranscriptomic reads.** Three previously sequenced soil  
569 metatranscriptomes collected from sampling locations matching three metagenomes (Extended  
570 Data Table S2) were downloaded from the Integrated Microbial Genomes database <sup>16</sup> and the  
571 NCBI Sequence Read Archive (SRA) <sup>17</sup>. These comprised: Kellogg Biological Station,  
572 Michigan, USA (Agricultural Land – Switchgrass, KBS) <sup>90</sup>; Anza Borrego Desert, California,  
573 USA (Hot Desert, ABD) <sup>92</sup>; Department of Meta, Colombia (Tropical Peatland, DMC) <sup>93</sup>. Raw  
574 metatranscriptomes were quality filtered as described above and SortMeRNA (version 4.2.0) <sup>100</sup>  
575 was used to remove ribosomal RNA (rRNA) reads. To determine transcript abundances of 43  
576 metabolic marker genes described above, filtered reads with lengths of at least 100 bp were  
577 searched for their presence using DIAMOND *blastx* and read counts were normalized to reads  
578 per kilobase per million (RPKM). To calculate transcript abundances of genes annotated from  
579 metagenomic assemblies and MAGs, Salmon v1.3.0 <sup>101</sup> was used in mapping-based mode  
580 (salmon quant, --validate Mappings) and read counts were normalised to transcripts per million  
581 (TPM). For genes predicted from metagenomic assemblies, they were firstly clustered at 95%  
582 sequence identity to produce a non-redundant gene catalog as a ‘reference transcriptome’ using  
583 CD-HIT v4.8.1 <sup>102</sup> before Salmon quantifications.

584

585 **Phylogenetic analysis.** Phylogenetic trees were constructed to understand the distribution and  
586 diversity of bacteria and archaea consuming inorganic energy sources. Trees were constructed  
587 for the group 1 and 2 [NiFe]-hydrogenase large subunits, group 3 [NiFe]-hydrogenase large  
588 subunits, CoxL, PmoA, RbcL, AmoA, NxrA, Sqr, SoxB, and DsrA. In all cases, protein  
589 sequences retrieved from the MAGs reads by homology-based searches were aligned against a  
590 subset of reference sequences from the custom protein databases using ClustalW in MEGA7 <sup>103</sup>.  
591 Evolutionary relationships were visualized by constructing a maximum-likelihood phylogenetic  
592 tree; specifically, initial trees for the heuristic search were obtained automatically by applying  
593 Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT



594 model, and then selecting the topology with superior log likelihood value. All residues were used  
595 and trees were bootstrapped with 50 replicates. To further visualize the diversity of trace gas  
596 oxidizers, neighbor-joining trees were constructed using the binned and unbinned sequences for  
597 group 1 and 2 [NiFe]-hydrogenase large subunits and CoxL; these trees were constructed using  
598 the Poisson model with gaps treated with pairwise deletion and bootstrapped with 50 replicates.

599

600 **Protein homology modelling.** Structural homology models for particulate methane  
601 monooxygenase subunits PmoA, PmoB, and PmoC from *Candidatus Methylophilum kingii*,  
602 *Methylacidiphilum infernorum* (PmoA: WP\_012463843, PmoB: WP\_012463845, PmoC:  
603 WP\_012463844) and *Candidatus Methylophilum limnetica* (PmoA: WP\_107560874, PmoB:  
604 WP\_107560875, PmoC: WP\_107560873) were generated using the Phyre2 webserver<sup>104</sup>. Direct  
605 threading of Pmo subunit amino acid sequences was performed using the closest structurally  
606 characterized particulate methane monooxygenase homologue to each strain, namely those of  
607 *Methylocystis* sp. strain M (PDB ID: 3RFR)<sup>105</sup> for *Ca. M. kingii* and *Methylococcus capsulatus*  
608 (PDB ID: 3RGB)<sup>105</sup> for *M. infernorum* and *Ca. M. limnetica*. Direct threading was used to avoid  
609 modelling using earlier erroneously modelled Pmo structures (PDB IDs: 1YEW and 3CHX)  
610 present in the PDB.

611

612 **Analysis of community composition and diversity.** Bacterial and archaeal community  
613 composition was determined from the pre-processed metagenomic reads with SingleM v0.12.1  
614 (<https://github.com/wwood/singlem>). In total, 28 HMM searches were performed against 14  
615 single-copy ribosomal marker genes. The gene for single-copy ribosomal protein L16/L10E  
616 (*rplP*) was selected for downstream analysis and sequences were clustered *de novo* into  
617 operational taxonomic units at a sequence identity threshold of 97%. Taxonomic assignment was  
618 carried out using the Genome Taxonomy Database<sup>84</sup>. Community richness and beta diversity  
619 were calculated using the phyloseq<sup>106</sup> and R package VEGAN<sup>107</sup>. To account for differences in  
620 richness between samples, all sequences were rarefied to within 90% of the minimum sequence  
621 count. Observed richness and estimated richness (Chao1) were calculated. First, beta diversity  
622 (Bray-Curtis) was calculated and visualised using a multidimensional scaling ordination (MDS),  
623 then significance testing was carried out using a permutational analysis of variance  
624 (PERMANOVA) to test for significant differences in community structure between ecosystems  
625 and between depth profiles. A beta dispersion test (PERMDISP) was used to ascertain if  
626 significant differences in community structure were due to data dispersion.

627

628 **Thermodynamic modelling.** Cell-specific power (i.e. Gibbs energy per unit time per microbial  
629 cell),  $P$  (W), was calculated for each reaction using measured *ex situ* oxidation rates according to:  
630

$$631 \quad P = \frac{r \cdot \Delta G_r}{B} \quad (\text{EQ. 4})$$

632  
633 where  $r$  denotes the rate of reaction ( $\text{mol s}^{-1} \text{g}_{\text{dry soil}}^{-1}$ ),  $\Delta G_r$  represents the Gibbs energy of the  
634 reaction ( $\text{J mol}^{-1}$ ), and  $B$  ( $\text{cells g}_{\text{dry soil}}^{-1}$ ) is the number of microbial cells carrying out each of the  
635 following the reactions:  $\text{H}_2 + 0.5 \text{O}_2 \rightarrow \text{H}_2\text{O}$  (dihydrogen oxidation);  $\text{CO} + 0.5 \text{O}_2 \rightarrow \text{CO}_2$   
636 (carbon monoxide oxidation); and  $\text{CH}_4 + 2 \text{O}_2 \rightarrow \text{CO}_2 + 2 \text{H}_2\text{O}$  (methane oxidation). Values of  
637  $\Delta G_r$  are calculated using:

$$638 \quad \Delta G_r = \Delta G_r^0 + RT \ln Q_r \quad (\text{EQ. 5})$$

639  
640  
641 where  $\Delta G_r^0$  and  $Q_r$  refer to the standard Gibbs energy and the reaction quotient of the indicated  
642 reaction, respectively,  $R$  represents the gas constant, and  $T$  denotes temperature in Kelvin. Values  
643 of  $\Delta G_r^0$  were calculated using the revised-HKF equations of state<sup>108–110</sup>, the SUPCRT92  
644 software package<sup>111</sup>, and thermodynamic data taken from references<sup>109,112–115</sup>.

645  
646 Values of  $Q_r$  are calculated for each reaction using:

$$647 \quad Q_r = \prod a_g^{v_i} \quad (\text{EQ. 6})$$

648  
649  
650 where  $a_i$  stands for the activity of the  $i^{\text{th}}$  species and  $v_i$  corresponds to the stoichiometric  
651 coefficient of the  $i^{\text{th}}$  species in the reaction of interest. Activities of gas-phase compounds,  $a_g$ ,  
652 were calculated using:

$$653 \quad a_g = \frac{f_g}{f_g^0} \quad (\text{EQ. 7})$$

654  
655 where  $f_g$  and  $f_g^0$  designate the fugacity and standard state fugacity of the respective gas. Due to  
656 the low temperatures and pressures of soil ecosystems, fugacity coefficients for all gases are one  
657 (see<sup>116</sup>). Therefore, partial pressures are equivalent to fugacity since  $f_g^0$  was taken to be 1 bar.  
658 Gibbs energy calculations were carried out at 20°C and 1 bar. The concentrations of reactants in  
659 the catabolic reactions and reaction rates were measured by gas chromatography as described  
660 above.

661

662 Estimates of the number of microbial cells carrying out each reaction ( $B$ ) were obtained by  
663 calculating the proportion of cells performing a specific catabolic reaction in our 16S rRNA gene  
664 copy number dataset, assuming all cells to be active. The proportion of a specific functional  
665 community was retrieved from the relative abundance of the individual biomarker genes  
666 (encoding the large subunits of group 1c, 1d, 1f, 1h, and 2a [NiFe]-hydrogenase, carbon  
667 monoxide dehydrogenase (CoxL), particulate methane monooxygenase (PmoA), and soluble  
668 methane monooxygenase (MmoX)) in our metagenomics dataset. An average of 4.2 16S rRNA  
669 copies per cell was assumed for cell number estimation<sup>117</sup>.

670

671 To compare our approach to previous literature, we repeated an earlier attempt to estimate the  
672 theoretical maximum population of trace gas oxidizing bacteria from modelled  $\Delta G_r$  and measured  
673 reaction rates<sup>68</sup>:

674

$$675 \quad N_t = \frac{N_c}{m_E} \Delta G_r r \quad (\text{EQ. 8})$$

676

677 Here,  $m_E$  is a mean maintenance energy at 293K in [kJ (C-mol biomass)<sup>-1</sup> h<sup>-1</sup>], estimated from  
678 pure-culture data of different metabolisms<sup>63</sup>;  $N_c$  is the number of cells per C-mol of biomass ( $1.4$   
679  $\times 10^{14}$  cells C-mol<sup>-1</sup><sup>68</sup>, equivalent to  $8.6 \times 10^{-14}$  g C dry cell<sup>-1</sup>); and  $r$  is the trace-gas reaction rate  
680 as defined above. The ratio  $N_t / B$  (the fraction of the estimated trace gas oxidizing population that  
681 can be supported given a specific maintenance energy) is equivalent to dividing cell-specific  
682 power by the given maintenance energy, after conversion to W cell<sup>-1</sup> via  $N_c$ :

683

$$684 \quad \frac{N_t}{B} = \frac{\frac{N_c \Delta G_r r}{m_E}}{B} = \frac{P}{\frac{m_E}{N_c}} \quad (\text{EQ. 9})$$

685

686

687 **Reporting summary.** Further information on research design is available in the Nature Research  
688 Reporting Summary linked to this article.

## 689 **Data Availability**

690 All data supporting the findings of the present study are available. All numerical data used to  
691 make figures is provided in **Source Data 1** (XLSX format) and all raw phylogenetic trees are  
692 provided in Newick (NWK) format in **Source Data 2** (ZIP format). All metagenomes sequenced  
693 for this project can be accessed at the Sequence Read Archive with accession number  
694 PRJNA656125 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA656125>) and all metagenome-  
695 assembled genomes are available at FigShare  
696 ([https://figshare.com/articles/dataset/Metagenome\\_Assembled\\_Genomes/12782543](https://figshare.com/articles/dataset/Metagenome_Assembled_Genomes/12782543)). All  
697 previously sequenced metagenomes and metatranscriptomes analysed in this study are available  
698 at NCBI BioProject with the accession numbers listed in **Supplementary Table 2**.

699

## 700 **References**

701

- 702 1. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome.  
703 *Nat. Rev. Microbiol.* **15**, 579 (2017).
- 704 2. Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L. & Lilley, A. K. The contribution  
705 of species richness and composition to bacterial services. *Nature* **436**, 1157 (2005).
- 706 3. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil.  
707 *Science* **359**, 320–325 (2018).
- 708 4. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**,  
709 233–237 (2018).
- 710 5. Jones, S. E. & Lennon, J. T. Dormancy contributes to the maintenance of microbial  
711 diversity. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5881–5886 (2010).
- 712 6. Lennon, J. T. & Jones, S. E. Microbial seed banks: the ecological and evolutionary  
713 implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 (2011).
- 714 7. Schimel, J. & Schaeffer, S. M. Microbial control over carbon cycling in soil. *Front.*  
715 *Microbiol.* **3**, 348 (2012).
- 716 8. Janssen, P. H. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and  
717 16S rRNA genes. *Appl. Environ. Microbiol.* **72**, 1719–1728 (2006).
- 718 9. Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A. & Janssen, P. H. Laboratory  
719 cultivation of widespread and previously uncultured soil bacteria. *Appl. Environ.*  
720 *Microbiol.* **69**, 7210–7215 (2003).
- 721 10. Conrad, R. Soil microorganisms as controllers of atmospheric trace gases (H<sub>2</sub>, CO, CH<sub>4</sub>,  
722 OCS, N<sub>2</sub>O, and NO). *Microbiol. Rev.* **60**, 609–640 (1996).
- 723 11. Beekman, F., Motte, H. & Beekman, T. Nitrification in agricultural soils: impact, actors  
724 and mitigation. *Curr. Opin. Biotechnol.* **50**, 166–173 (2018).

- 725 12. Khdhiri, M. *et al.* Soil carbon content and relative abundance of high affinity H<sub>2</sub>-oxidizing  
726 bacteria predict atmospheric H<sub>2</sub> soil uptake activity better than soil microbial community  
727 composition. *Soil Biol. Biochem.* **85**, 1–9 (2015).
- 728 13. Ehhalt, D. H. & Rohrer, F. The tropospheric cycle of H<sub>2</sub>: a critical review. *Tellus B* **61**,  
729 500–535 (2009).
- 730 14. Novelli, P. C., Masarie, K. A., Tans, P. P. & Lang, P. M. Recent changes in atmospheric  
731 carbon monoxide. *Science* **263**, 1587–1590 (1994).
- 732 15. Novelli, P. C., Masarie, K. A. & Lang, P. M. Distributions and recent changes of carbon  
733 monoxide in the lower troposphere. *J. Geophys. Res. Atmos.* **103**, 19015–19033 (1998).
- 734 16. Kirschke, S. *et al.* Three decades of global methane sources and sinks. *Nat. Geosci* **6**,  
735 813–823 (2013).
- 736 17. Constant, P., Poissant, L. & Villemur, R. Tropospheric H<sub>2</sub> budget and the response of its  
737 soil uptake under the changing environment. *Sci. Total Environ.* **407**, 1809–1823 (2009).
- 738 18. Berney, M., Greening, C., Conrad, R., Jacobs, W. R. & Cook, G. M. An obligately aerobic  
739 soil bacterium activates fermentative hydrogen production to survive reductive stress  
740 during hypoxia. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11479–11484 (2014).
- 741 19. Tarr, M. A., Miller, W. L. & Zepp, R. G. Direct carbon monoxide photoproduction from  
742 plant matter. *J. Geophys. Res. Atmos.* **100**, 11403–11413 (1995).
- 743 20. King, G. M. & Weber, C. F. Distribution, diversity and ecology of aerobic CO-oxidizing  
744 bacteria. *Nat. Rev. Microbiol.* **5**, 107–118 (2007).
- 745 21. Conrad, R. The global methane cycle: recent advances in understanding the microbial  
746 processes involved. *Environ. Microbiol. Rep.* **1**, 285–292 (2009).
- 747 22. Constant, P., Chowdhury, S. P., Pratscher, J. & Conrad, R. Streptomycetes contributing to  
748 atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-  
749 affinity [NiFe]-hydrogenase. *Environ. Microbiol.* **12**, 821–829 (2010).
- 750 23. Greening, C., Berney, M., Hards, K., Cook, G. M. & Conrad, R. A soil actinobacterium  
751 scavenges atmospheric H<sub>2</sub> using two membrane-associated, oxygen-dependent [NiFe]  
752 hydrogenases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4257–4261 (2014).
- 753 24. Cordero, P. R. F. *et al.* Atmospheric carbon monoxide oxidation is a widespread  
754 mechanism supporting microbial survival. *ISME J.* **13**, 2868–2881 (2019).
- 755 25. Mörsdorf, G., Frunzke, K., Gadkari, D. & Meyer, O. Microbial growth on carbon  
756 monoxide. *Biodegradation* **3**, 61–82 (1992).
- 757 26. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution  
758 indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *ISME J.*  
759 **10**, 761–777 (2016).
- 760 27. Tveit, A. T. *et al.* Widespread soil bacterium that oxidizes atmospheric methane. *Proc.*  
761 *Natl. Acad. Sci. U. S. A.* **116**, 8515–8524 (2019).
- 762 28. Islam, Z. F. *et al.* A widely distributed hydrogenase oxidises atmospheric H<sub>2</sub> during

- 763 bacterial growth. *ISME J.* 10.1038/s41396-020-0713-4 (2020).
- 764 29. King, G. M. Molecular and culture-based analyses of aerobic carbon monoxide oxidizer  
765 diversity. *Appl. Environ. Microbiol.* **69**, 7257–7265 (2003).
- 766 30. Greening, C. *et al.* Persistence of the dominant soil phylum Acidobacteria by trace gas  
767 scavenging. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10497–10502 (2015).
- 768 31. Berney, M. & Cook, G. M. Unique flexibility in energy metabolism allows mycobacteria  
769 to combat starvation and hypoxia. *PLoS One* **5**, e8614 (2010).
- 770 32. Greening, C., Villas-Bôas, S. G., Robson, J. R., Berney, M. & Cook, G. M. The growth  
771 and survival of *Mycobacterium smegmatis* is enhanced by co-metabolism of atmospheric  
772 H<sub>2</sub>. *PLoS One* **9**, e103034 (2014).
- 773 33. Liot, Q. & Constant, P. Breathing air to save energy – new insights into the  
774 ecophysiological role of high-affinity [NiFe]-hydrogenase in *Streptomyces avermitilis*.  
775 *Microbiologyopen* **5**, 47–59 (2016).
- 776 34. Greening, C., Grinter, R. & Chiri, E. Uncovering the metabolic strategies of the dormant  
777 microbial majority: towards integrative approaches. *mSystems* **4**, e00107-19 (2019).
- 778 35. Khdhiri, M., Piché-Choquette, S., Tremblay, J., Tringe, S. G. & Constant, P. Meta-omics  
779 survey of [NiFe]-hydrogenase genes fails to capture drastic variations in H<sub>2</sub>-oxidation  
780 activity measured in three soils exposed to H<sub>2</sub>. *Soil Biol. Biochem.* **125**, 239–243 (2018).
- 781 36. Quiza, L., Lalonde, I., Guertin, C. & Constant, P. Land-use influences the distribution and  
782 activity of high affinity CO-oxidizing bacteria Associated to type I-coxL genotype in soil.  
783 *Front. Microbiol.* **5**, 271 (2014).
- 784 37. Ji, M. *et al.* Atmospheric trace gases support primary production in Antarctic desert  
785 surface soil. *Nature* **552**, 400–403 (2017).
- 786 38. Lynch, R. C., Darcy, J. L., Kane, N. C., Nemergut, D. R. & Schmidt, S. K. Metagenomic  
787 evidence for metabolism of trace atmospheric gases by high-elevation desert  
788 Actinobacteria. *Front. Microbiol.* **5**, 698 (2014).
- 789 39. King, G. M. Contributions of atmospheric CO and hydrogen uptake to microbial dynamics  
790 on recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* **69**, 4067–4075 (2003).
- 791 40. Cordero, P. R. F. *et al.* Carbon monoxide dehydrogenases enhance bacterial survival by  
792 respiring atmospheric CO. *bioRxiv* 628081 (2019) doi:10.1101/628081.
- 793 41. Constant, P., Chowdhury, S. P., Hesse, L., Pratscher, J. & Conrad, R. Genome data mining  
794 and soil survey for the novel Group 5 [NiFe]-hydrogenase to explore the diversity and  
795 ecological importance of presumptive high-affinity H<sub>2</sub>-oxidizing bacteria. *Appl. Environ.*  
796 *Microbiol.* **77**, 6027–6035 (2011).
- 797 42. Pandelia, M. E., Lubitz, W. & Nitschke, W. Evolution and diversification of Group 1  
798 [NiFe] hydrogenases. Is there a phylogenetic marker for O<sub>2</sub>-tolerance? *Biochim. Biophys.*  
799 *Acta - Bioenerg.* **1817**, 1565–1575 (2012).
- 800 43. Myers, M. R. & King, G. M. Isolation and characterization of *Acidobacterium ailaau* sp.  
801 nov., a novel member of Acidobacteria subdivision 1, from a geothermally heated

- 802 Hawaiian microbial mat. *Int. J. Syst. Evol. Microbiol.* **66**, 5328–5335 (2016).
- 803 44. Cunliffe, M. Correlating carbon monoxide oxidation with *cox* genes in the abundant  
804 marine *Roseobacter* clade. *ISME J.* **5**, 685 (2011).
- 805 45. van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature*  
806 **528**, 555 (2015).
- 807 46. Daims, H. *et al.* Complete nitrification by *Nitrospira* bacteria. *Nature* **528**, 504 (2015).
- 808 47. Knief, C. Diversity and habitat preferences of cultivated and uncultivated aerobic  
809 methanotrophic bacteria evaluated based on *pmoA* as molecular marker. *Front. Microbiol.*  
810 **6**, 1346 (2015).
- 811 48. Chiri, E. *et al.* Termite mounds contain soil-derived methanotroph communities  
812 kinetically adapted to elevated methane concentrations. *ISME J.* 1–17 (2020).
- 813 49. Dunfield, P. F. *et al.* Methane oxidation by an extremely acidophilic bacterium of the  
814 phylum Verrucomicrobia. *Nature* **450**, 879–882 (2007).
- 815 50. Haroon, M. F. *et al.* Anaerobic oxidation of methane coupled to nitrate reduction in a  
816 novel archaeal lineage. *Nature* **500**, 567–70 (2013).
- 817 51. Khadka, R. *et al.* Evolutionary history of copper membrane monooxygenases. *Front.*  
818 *Microbiol.* **9**, 2493 (2018).
- 819 52. Van Asperen, H. *et al.* The role of photo-and thermal degradation for CO<sub>2</sub> and CO fluxes  
820 in an arid ecosystem. *Biogeosciences* **12**, (2015).
- 821 53. Moxley, J. M. & Smith, K. A. Carbon monoxide production and emission by some  
822 Scottish soils. *Tellus B Chem. Phys. Meteorol.* **50**, 151–162 (1998).
- 823 54. Dutaur, L. & Verchot, L. V. A global inventory of the soil CH<sub>4</sub> sink. *Global Biogeochem.*  
824 *Cycles* **21**, GB4013 (2007).
- 825 55. McLearn, N. & Dong, Z. Microbial nature of the hydrogen-oxidizing agent in hydrogen-  
826 treated soil. *Biol. Fertil. Soils* **35**, 465–469 (2002).
- 827 56. King, G. M. Attributes of atmospheric carbon monoxide oxidation by Maine forest soils.  
828 *Appl. Environ. Microbiol.* **65**, 5257–5264 (1999).
- 829 57. Kolb, S., Knief, C., Dunfield, P. F. & Conrad, R. Abundance and activity of uncultured  
830 methanotrophic bacteria involved in the consumption of atmospheric methane in two  
831 forest soils. *Environ. Microbiol.* **7**, 1150–1161 (2005).
- 832 58. Priemé, A. & Christensen, S. Seasonal and spatial variation of methane oxidation in a  
833 Danish spruce forest. *Soil Biol. Biochem.* **29**, 1165–1172 (1997).
- 834 59. Chen, Q., Popa, M. E., Batenburg, A. M. & Röckmann, T. Isotopic signatures of  
835 production and uptake of H<sub>2</sub> by soil. *Atmos. Chem. Phys.* **15**, 13003–13021 (2015).
- 836 60. DeLong, J. P., Okie, J. G., Moses, M. E., Sibly, R. M. & Brown, J. H. Shifts in metabolic  
837 scaling, production, and efficiency across major evolutionary transitions of life. *Proc.*  
838 *Natl. Acad. Sci.* **107**, 12941–12945 (2010).

- 839 61. Kempes, C. P. *et al.* Drivers of bacterial maintenance and minimal energy requirements.  
840 *Front. Microbiol.* **8**, 31 (2017).
- 841 62. Marschall, E., Jogler, M., Henßge, U. & Overmann, J. Large-scale distribution and  
842 activity patterns of an extremely low-light-adapted population of green sulfur bacteria in  
843 the Black Sea. *Environ. Microbiol.* **12**, 1348–1362 (2010).
- 844 63. Tijhuis, L., Van Loosdrecht, M. C. & Heijnen, J. J. A thermodynamically based  
845 correlation for maintenance gibbs energy requirements in aerobic and anaerobic  
846 chemotrophic growth. *Biotechnol. Bioeng.* **42**, 509–519 (1993).
- 847 64. LaRowe, D. E. & Amend, J. P. Power limits for microbial life. *Front. Microbiol.* **6**, 718  
848 (2015).
- 849 65. Bradley, J. A. *et al.* Widespread energy limitation to life in global seafloor sediments.  
850 *Sci. Adv.* **6**, eaba0697 (2020).
- 851 66. Røy, H. *et al.* Aerobic microbial respiration in 86-million-year-old deep-sea red clay.  
852 *Science* **336**, 922–925 (2012).
- 853 67. Conrad, R. Capacity of aerobic microorganisms to utilize and grow on atmospheric trace  
854 gases (H<sub>2</sub>, CO, CH<sub>4</sub>). in *Current Perspectives in Microbial Ecology* (eds. Klug, M. G. &  
855 Reddy, C. A.) 461–467 (American Society for Microbiology, 1984).
- 856 68. Conrad, R. Soil microorganisms oxidizing atmospheric trace gases (CH<sub>4</sub>, CO, H<sub>2</sub>, NO).  
857 *Indian J. Microbiol.* **39**, 193–203 (1999).
- 858 69. Meredith, L. K. *et al.* Ecosystem fluxes of hydrogen in a mid-latitude forest driven by  
859 soil microorganisms and plants. *Glob. Chang. Biol.* **23**, 906–919 (2017).
- 860 70. Carini, P. *et al.* Relic DNA is abundant in soil and obscures estimates of soil microbial  
861 diversity. *Nat. Microbiol.* **2**, 1–6 (2016).
- 862 71. Schmitz, R. A. *et al.* The thermoacidophilic methanotroph *Methylacidiphilum*  
863 *fumariolicum* SolV oxidizes subatmospheric H<sub>2</sub> with a high-affinity, membrane-associated  
864 [NiFe] hydrogenase. *ISME J.* **14**, 1223–1232 (2020).
- 865 72. Islam, Z. F. *et al.* Two Chloroflexi classes independently evolved the ability to persist on  
866 atmospheric hydrogen and carbon monoxide. *ISME J.* **13**, 1801–1813 (2019).
- 867 73. Hüppi, R. *et al.* Restricting the nonlinearity parameter in soil greenhouse gas flux  
868 calculation for more reliable flux estimates. *PLoS One* **13**, (2018).
- 869 74. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of  
870 sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4516–4522 (2011).
- 871 75. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-  
872 resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
- 873 76. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-  
874 node solution for large and complex metagenomics assembly via succinct de Bruijn graph.  
875 *Bioinformatics* **31**, 1674–1676 (2015).
- 876 77. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile



- 877 metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
- 878 78. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately  
879 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165  
880 (2015).
- 881 79. Kang, D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient  
882 genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
- 883 80. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning  
884 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**,  
885 605–607 (2015).
- 886 81. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate  
887 genomic comparisons that enables improved genome recovery from metagenomes through  
888 de-replication. *ISME J.* **11**, 2864 (2017).
- 889 82. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:  
890 assessing the quality of microbial genomes recovered from isolates, single cells, and  
891 metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- 892 83. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG)  
893 and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat.*  
894 *Biotechnol.* **35**, 725–731 (2017).
- 895 84. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny  
896 substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
- 897 85. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to  
898 classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927  
899 (2020).
- 900 86. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative  
901 analysis system. *Nucleic Acids Research* vol. 40 D115–22 (2012).
- 902 87. Leinonen, R., Sugawara, H., Shumway, M. & Collaboration, I. N. S. D. The sequence read  
903 archive. *Nucleic Acids Res.* **39**, D19–D21 (2010).
- 904 88. Jansson, J. K. Next generation ecosystem experiment (NGEE) in the Arctic. (2013)  
905 doi:10.25585/1487743.
- 906 89. Constant, P. Metagenomic and metatranscriptomic analysis of soil biogeochemical  
907 processes sustained by interspecific transfer of molecular hydrogen. (2014)  
908 doi:10.25585/1487738.
- 909 90. Tiedje, J. M. Metagenomic analysis of the rhizosphere of three biofuel crops at the KBS  
910 intensive site. (2013) doi:10.25585/1488010.
- 911 91. Mohn, W. W. Metagenomic and metatranscriptomic analysis of forest soil communities  
912 across North America. (2012) doi:10.25585/1487719.
- 913 92. Kalyuzhnaya, M. G. Systems level insights into methane cycling in arid and semi-arid  
914 ecosystems via community metagenomics and metatranscriptomics. (2016)  
915 doi:10.25585/1488146.

- 916 93. Cadillo-Quiroz, H. Microbial composition and metagenomic functional potential across  
917 Tropical Peatlands: comparative evaluation and modeling of C decomposition to  
918 greenhouse gases. (2015) doi:10.25585/1488117.
- 919 94. Oliveira, R. Unravelling microbial communities associated with native plant species from  
920 P-impooverished soils of a global biodiversity hotspot. (2017) doi:10.25585/1488283.
- 921 95. Pett-Ridge, J. Microbial carbon transformations in wet tropical soils: effects of redox  
922 fluctuation. (2017) doi:10.25585/1488160.
- 923 96. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site  
924 identification. *BMC Bioinformatics* **11**, 119 (2010).
- 925 97. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using  
926 DIAMOND. *Nat. Methods* **12**, 59 (2014).
- 927 98. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected  
928 biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
- 929 99. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment  
930 search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- 931 100. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal  
932 RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
- 933 101. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast  
934 and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
- 935 102. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-  
936 generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
- 937 103. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis  
938 version 7.0 for bigger datasets. *Mol. Biol. Evol.* msw054 (2016).
- 939 104. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2  
940 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
- 941 105. Smith, S. M. *et al.* Crystal structure and characterization of particulate methane  
942 monooxygenase from *Methylocystis* species strain M. *Biochemistry* **50**, 10231–10240  
943 (2011).
- 944 106. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive  
945 analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
- 946 107. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**,  
947 927–930 (2003).
- 948 108. Helgeson, H., Kirkham, D. & Flowers, G. Theoretical prediction of the thermodynamic  
949 behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of  
950 activity coefficients, osmotic coefficients, and apparent molal and standard and relative  
951 partial molal properties to 600 d. *Am. J. Sci.* **281**, 1249–1516 (1981).
- 952 109. Shock, E. L. & Helgeson, H. C. Calculation of the thermodynamic and transport  
953 properties of aqueous species at high pressures and temperatures: Correlation algorithms

- 954 for ionic species and equation of state predictions to 5 kb and 1000°C. *Geochim.*  
955 *Cosmochim. Acta* **52**, 2009–2036 (1988).
- 956 110. Shock, E. L. (University of C. at B., Oelkers, E. H., Johnson, J. W., Sverjensky, D. A. &  
957 Helgeson, H. C. Calculation of the thermodynamic properties of aqueous Species at high  
958 pressures and temperatures. *J. Chem. Soc. Faraday Trans.* **88**, 803–826 (1992).
- 959 111. Johnson, J. W., Oelkers, E. H. & Helgeson, H. C. A software package for calculating the  
960 standard molal thermodynamic properties of minerals, gases, aqueous species and  
961 reactions from 1 to 5000 bars and 0° to 1000° C. *Comput. Geosci.* **18**, 899–947 (1992).
- 962 112. Shock, E. L., Helgeson, H. C. & Sverjensky, D. A. Calculation of the thermodynamic and  
963 transport properties of aqueous species at high pressures and temperatures: Standard  
964 partial molal properties of inorganic neutral species. *Geochim. Cosmochim. Acta* **53**,  
965 2157–2183 (1989).
- 966 113. Shock, E. L. & Helgeson, H. C. Calculation of the thermodynamic and transport  
967 properties of aqueous species at high pressures and temperatures: Standard partial molal  
968 properties of organic species. *Geochim. Cosmochim. Acta* **54**, 915–945 (1990).
- 969 114. Sverjensky, D. A., Shock, E. L. & Helgeson, H. C. Prediction of the thermodynamic  
970 properties of aqueous metal complexes to 1000°C and 5 kb. *Geochim. Cosmochim. Acta*  
971 **61**, 1359–1412 (1997).
- 972 115. Schulte, M. D., Shock, E. L. & Wood, R. H. The temperature dependence of the standard-  
973 state thermodynamic properties of aqueous nonelectrolytes. *Geochim. Cosmochim. Acta*  
974 **65**, 3919–3930 (2001).
- 975 116. Amend, J. P. & LaRowe, D. E. Minireview: demystifying microbial reaction energetics.  
976 *Environ. Microbiol.* **21**, 3539–3547 (2019).
- 977 117. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes  
978 and its consequences for bacterial community analyses. *PLoS One* **8**, e57923 (2013).
- 979
- 980

981 **Acknowledgements**

982

983 This research was supported by an Australian Research Council DECRA Fellowship  
984 (DE170100310; C.G.), a Discovery Project grant (DP180101762; C.G. and P.L.M.C.), a Swiss  
985 National Foundation Early Mobility Postdoctoral Fellowship (P2EZP3\_178421; E.C.), a NSFC  
986 grant (41906076; X.D.), a NERC grant (NE/T010967/1; J.B.), a Humboldt Foundation  
987 Fellowship (J.B.), Monash University PhD scholarships (S.K.B. and P.M.L.), and an NHMRC  
988 EL2 Fellowship (APP1178715; salary for C.G.). We thank Maria Chuvochina for etymological  
989 advice, Steven Chown, Sergio Morales, Matthew Stott, and Gregory Cook for formative  
990 discussions, and the four reviewers for their helpful suggestions.

991

992 **Author information**

993

994 **Contributions**

995 C.G. conceived and supervised this study. S.K.B., C.G., E.C., and P.A.N. designed experiments  
996 and analyzed data. Different authors were responsible for performing fieldwork (S.K.B., E.C.,  
997 P.M.L., C.G.), laboratory work (S.K.B., T.J., E.C.), metagenome assembly (X.D.), metagenome  
998 analysis (S.K.B., C.G., X.D., E.C.), phylogenetic analysis (C.G.), protein modelling (R.G.), and  
999 thermodynamic modelling (J.A.B., P.A.N., D.L., E.C., S.K.B.). S.K.A. and P.L.M.C. provided  
1000 logistical and theoretical support. C.G., S.K.B., and P.A.N. wrote the paper with input from all  
1001 authors.

1002

1003 **Corresponding authors**

1004 Correspondence to Chris Greening (chris.greening@monash.edu) and Eleonora Chiri  
1005 (eleonora.chiri@monash.edu).

1006

1007 **Ethics declarations**

1008

1009 **Competing interests**

1010 The authors declare no competing financial interests.

1011

1012 **Additional information**

1013

1014 **Etymological Information**

1015 *Candidatus* Methylophilum (Me.thy.lo.tro'pi.cum.) N.L. n. *methylum* (from  
1016 French *méthyle* back-formation from French *méthylène*, coined from Gr. n. *methu*, wine and Gr.  
1017 n. *hulê*, wood), the methyl group; N.L. pref. *methylo-*, pertaining to the methyl radical; L. masc.  
1018 adj. *tropicus* pertaining to tropical zone of the Earth; N.L. neut. n. *Methylophilum* referring to a  
1019 methyl-using bacterium from the tropics. *Candidatus* Methylophilum kingii (ki'ngi.i.) N.L.  
1020 gen. n. *kingii* in honor of Prof. Gary King, who has pioneered research into the microbial  
1021 oxidation of methane, CO, and H<sub>2</sub>.  
1022

## 1023 **Figure legends**

1024

1025 **Fig. 1.** Energy conservation and carbon acquisition strategies of global soil bacteria. Homology-  
1026 based searches were used to detect marker genes for key metabolic processes in the metagenomic  
1027 short reads and metagenome-assembled genomes. **(a)** Heatmaps showing the abundance of each  
1028 gene in the metagenomic short reads of the four Australian depth profiles and eight global sites.  
1029 The percentage of the total community predicted to encode at least one of each gene for a  
1030 process is shown, based on normalization to single-copy marker genes. The genes detected are  
1031 usually present in single copies in genomes and, where genes performing similar functions are  
1032 collapsed together, the values are summed up to 100%. **(b)** Dot plot showing the metabolic  
1033 potential of the 757 metagenome-assembled genomes (MAGs). The size of each point represents  
1034 the number of genomes in each phylum that encode the gene of interest and the shading  
1035 represents the average genome completeness. Taxonomy assignment is based on the Genome  
1036 Taxonomy Database (GTDB)<sup>52</sup>.

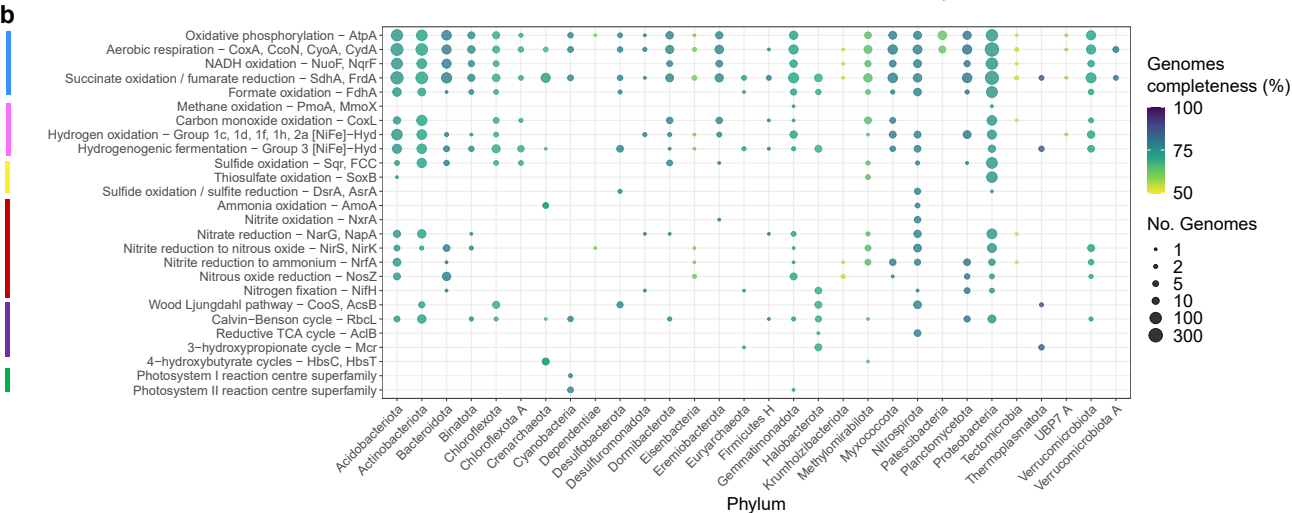
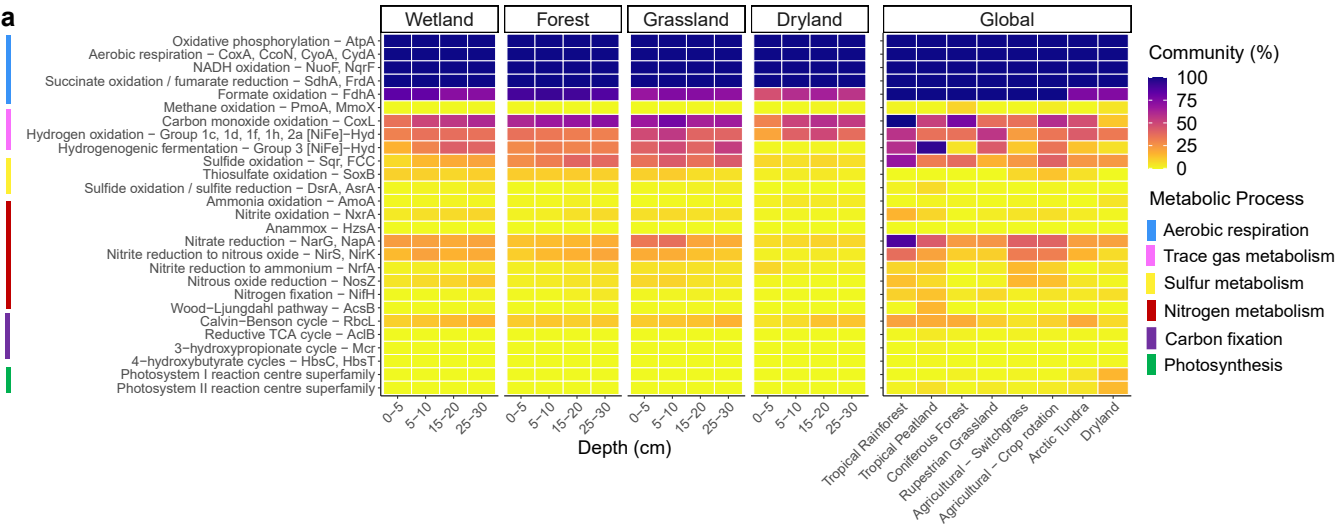
1037

1038 **Fig. 2.** Radial maximum-likelihood phylogenetic trees showing the sequence diversity and  
1039 taxonomic distribution of key enzymes associated with trace gas oxidation. The trees show the  
1040 amino acid sequences of **(a)** group 1 and 2 [NiFe]-hydrogenase large subunits and **(b)** form I  
1041 [MoCu]-carbon monoxide dehydrogenase large subunit (CoxL). The tree shows sequences from  
1042 the soil metagenome-assembled genomes (taxa names colored by phylum; 167 hydrogenase  
1043 sequences, 141 CO dehydrogenase sequences) alongside representative sequences from NCBI  
1044 reference genomes (taxa names shown in grey; 168 hydrogenase sequences, 171 CO  
1045 dehydrogenase sequences). The hydrogenase subgroups and carbon monoxide dehydrogenase  
1046 clades predicted to support trace gas oxidation in the sampled soils are highlighted. Both trees  
1047 were constructed using the JTT matrix-based model, used all sites, and were bootstrapped with  
1048 50 replicates. Source data is provided in Newick (nwk) format.

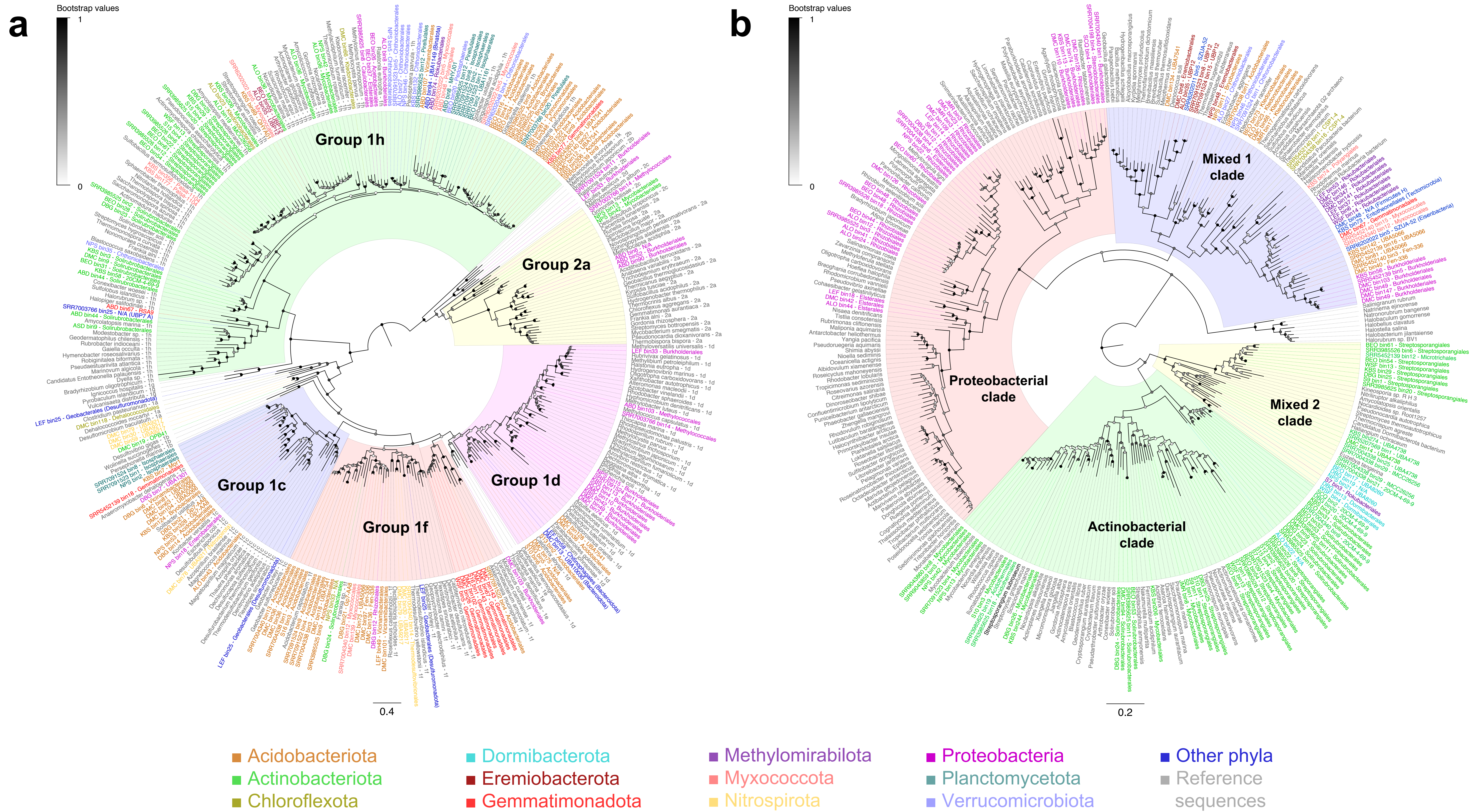
1049

1050 **Fig. 3.** Measurement of oxidation of the trace gases H<sub>2</sub>, CO, and CH<sub>4</sub> across four Australian soil  
1051 ecosystems. **(a)** Depth-resolved *in situ* gas concentrations from 0 to 16 cm depth at each  
1052 ecosystem (wetland, grassland, forest, dryland). **(b)** *In situ* soil-atmosphere gas fluxes ( $J_{\text{atm}}$ ;  
1053 positive values indicate net gas production, negative values indicate net gas consumption).  
1054 Dashed lines represent upper and lower simulated flux detection limits, with most CO and CH<sub>4</sub>  
1055 flux values falling within them. For **a** and **b**, measurements were performed in four separate soils  
1056 per biome and error bars showing standard deviations. **(c)** Depth-resolved, cell-specific *ex situ*  
1057 oxidation rates for each ecosystem of core samples from four different depths (0-5, 5-10, 15-20,

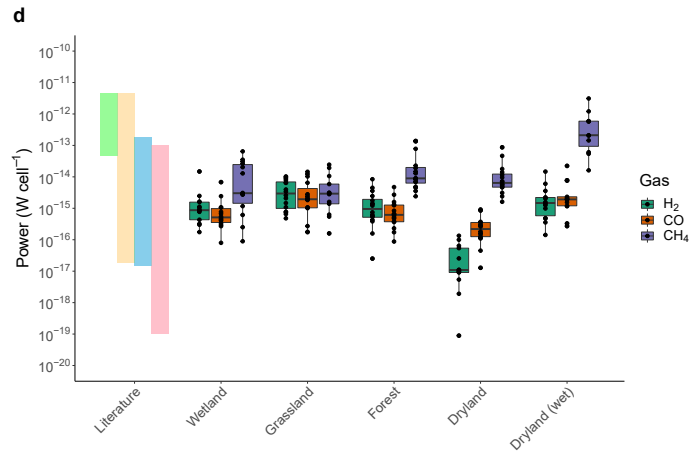
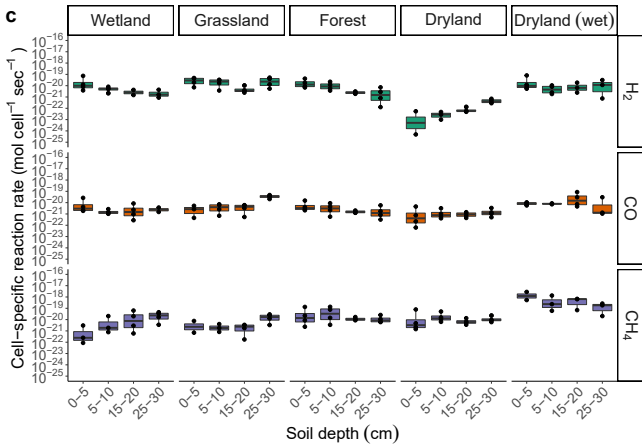
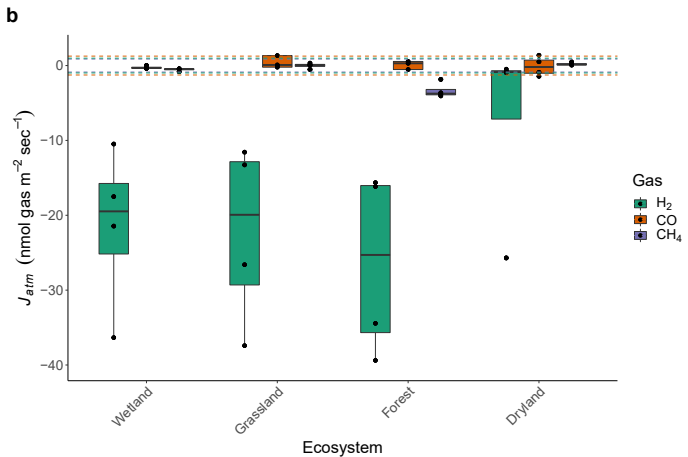
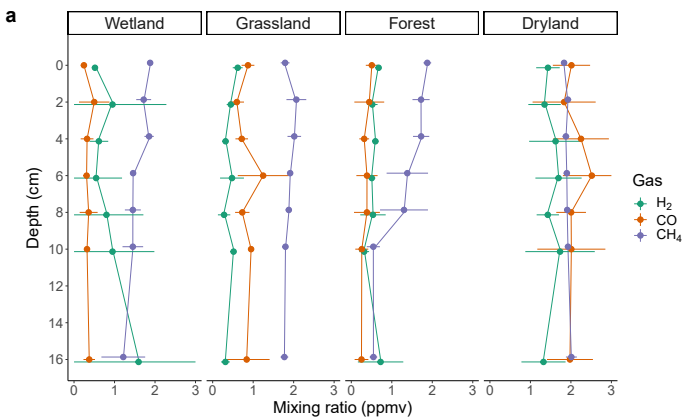
1058 25-30 cm). Measurements were performed with four biological replicates for each soil ecosystem  
1059 and treatment (dryland wet). Oxidation rate values are normalized to total predicted cell number  
1060 of trace gas oxidizers based on 16S rRNA gene qPCR and percentage of the community  
1061 predicted to oxidize each gas. **(d)** Amount of power per cell derived from the oxidation of each  
1062 trace gas. These were calculated using thermodynamic models based on *ex situ* rates measured  
1063 across the four depths per habitat. The colored bars depict the range of literature values of  
1064 maintenance energy requirements or endogenous metabolic rates of different pure cultures (green  
1065 <sup>63</sup>, yellow <sup>64</sup>, blue <sup>60</sup>) and hydrogen oxidizers in deep marine sediments (pink <sup>64,65</sup>). For **c** and **d**,  
1066 boxplots show means, lower and upper quartile, and minimum and maximum values.  
1067

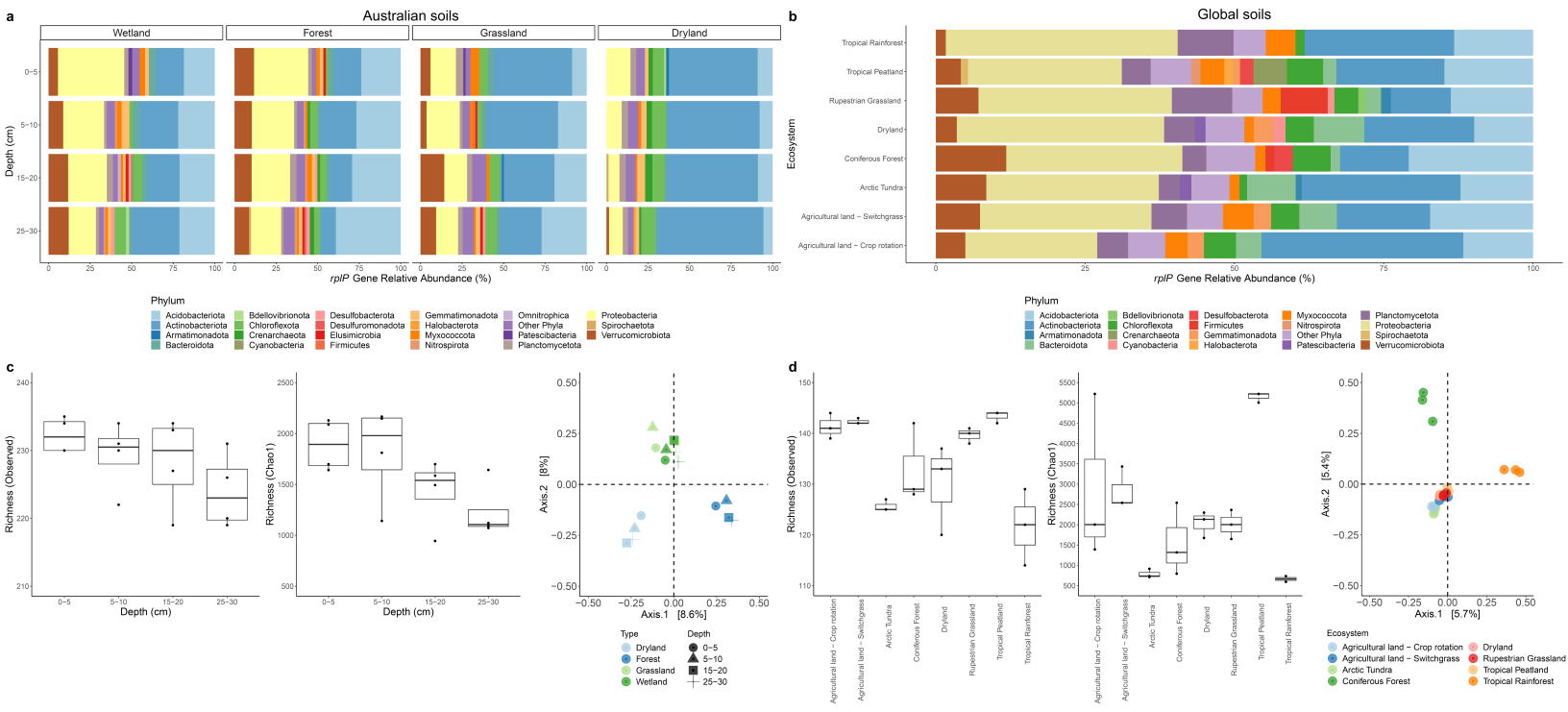


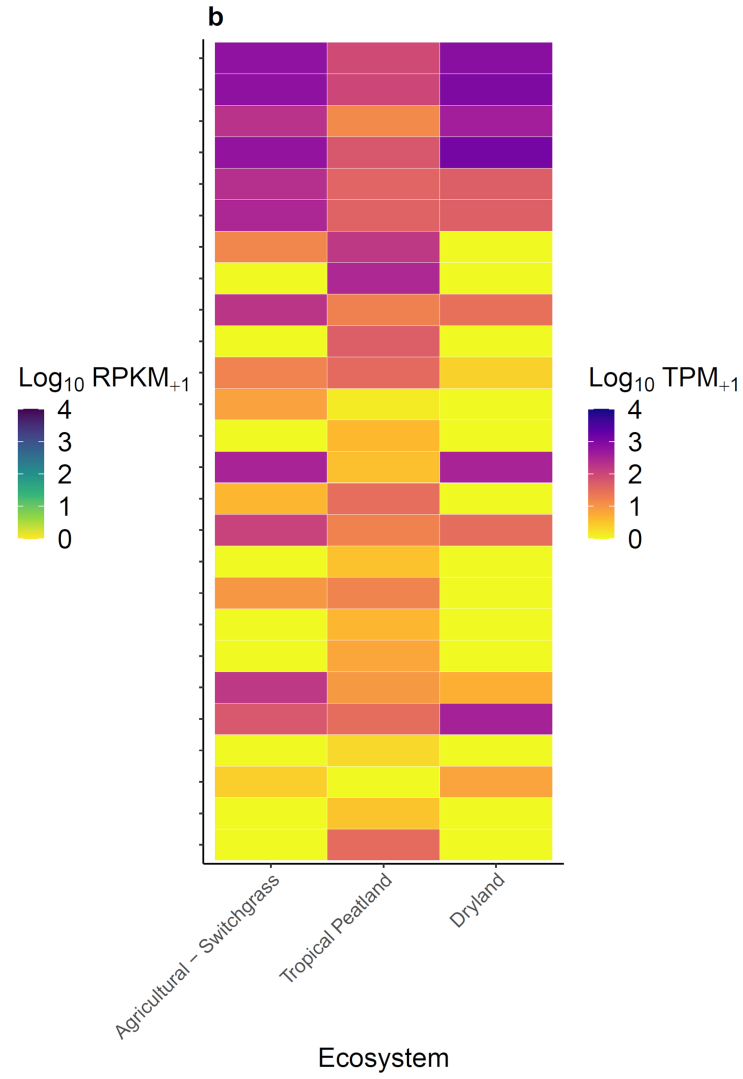
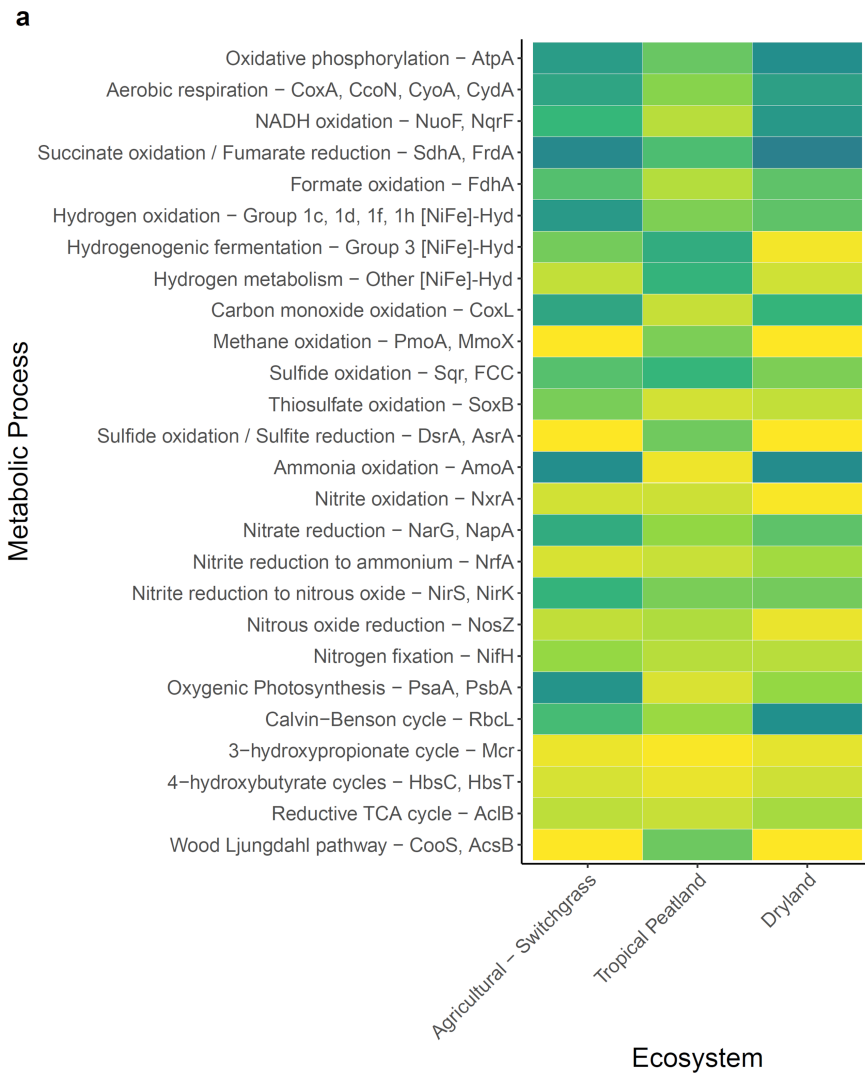


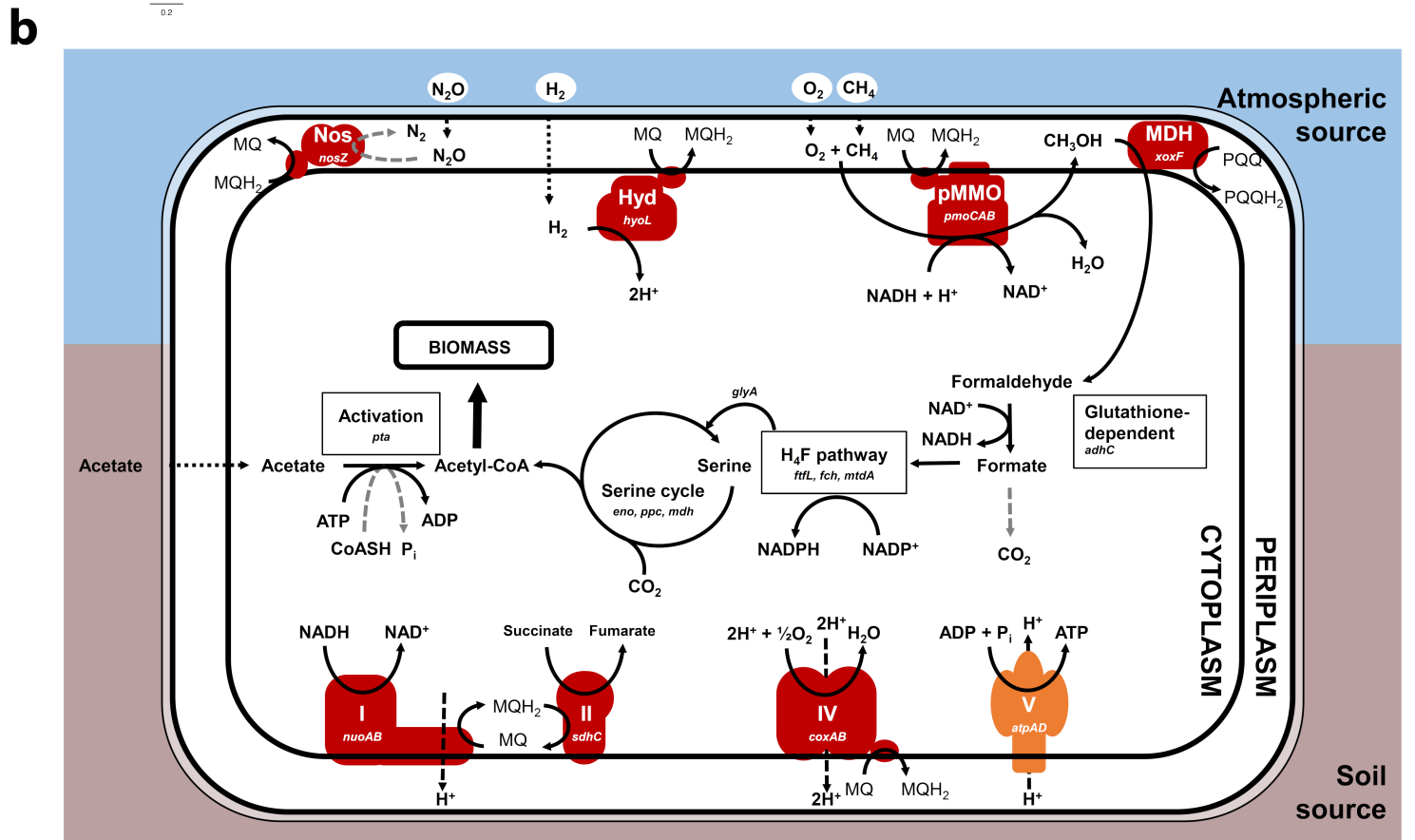
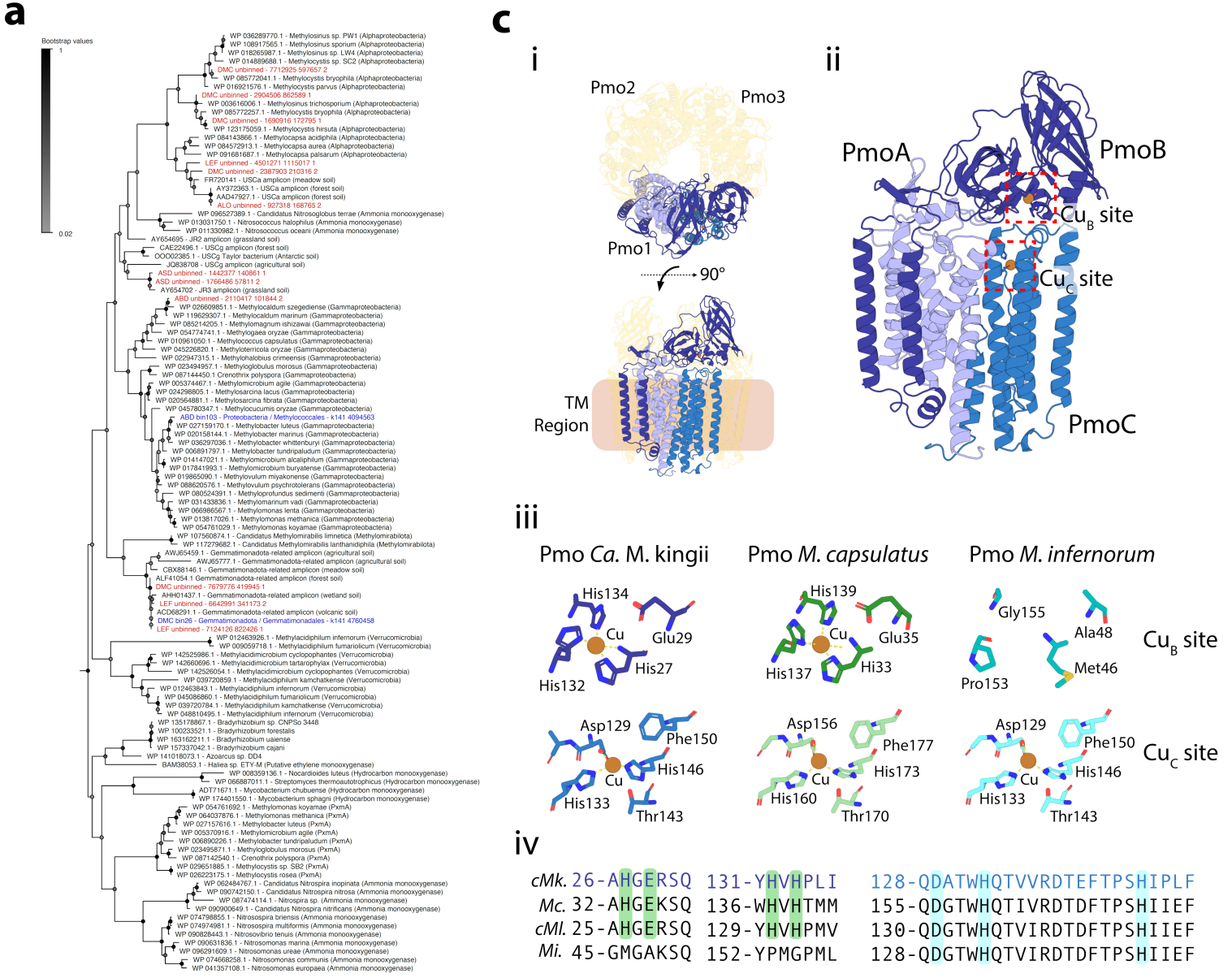


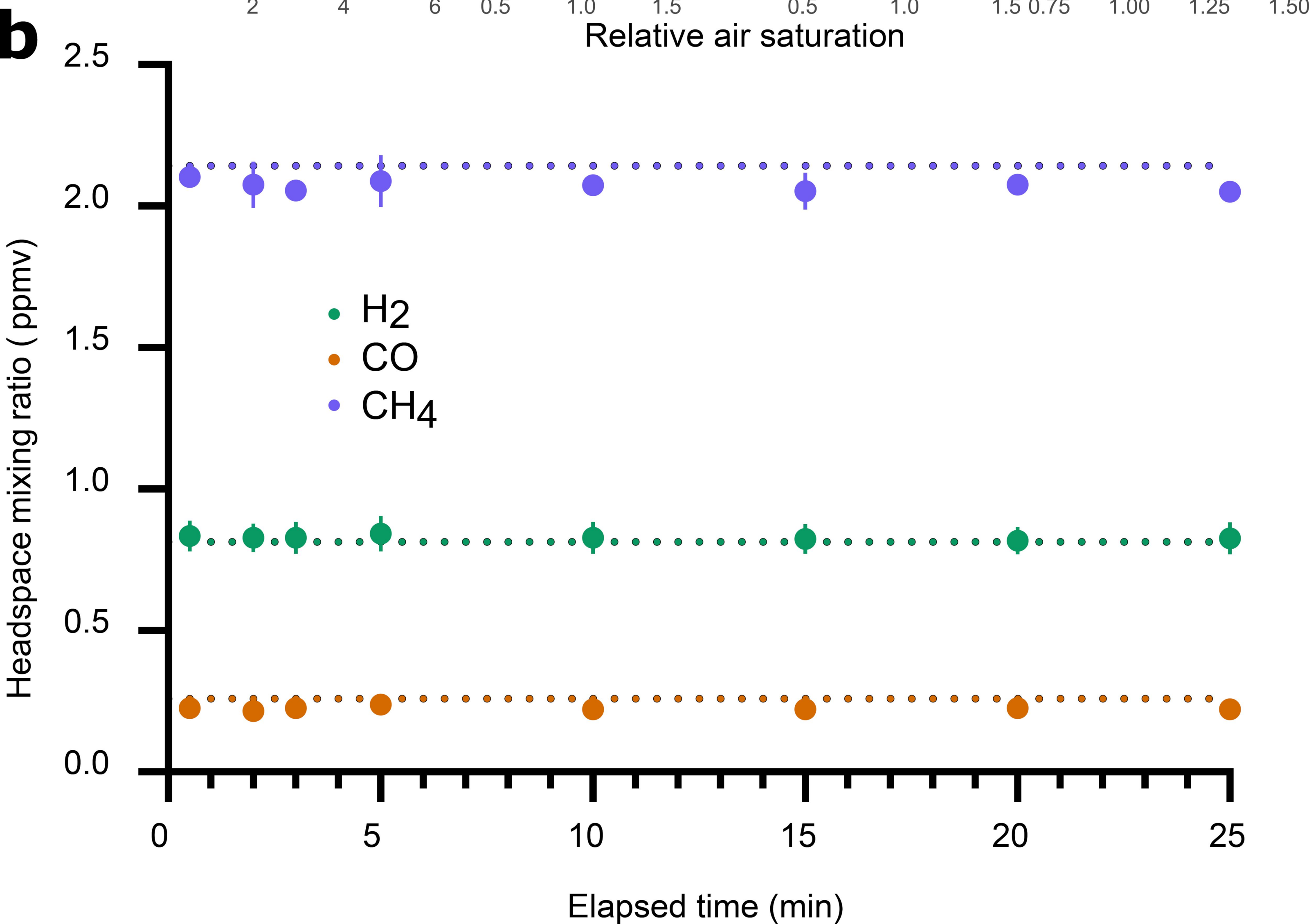
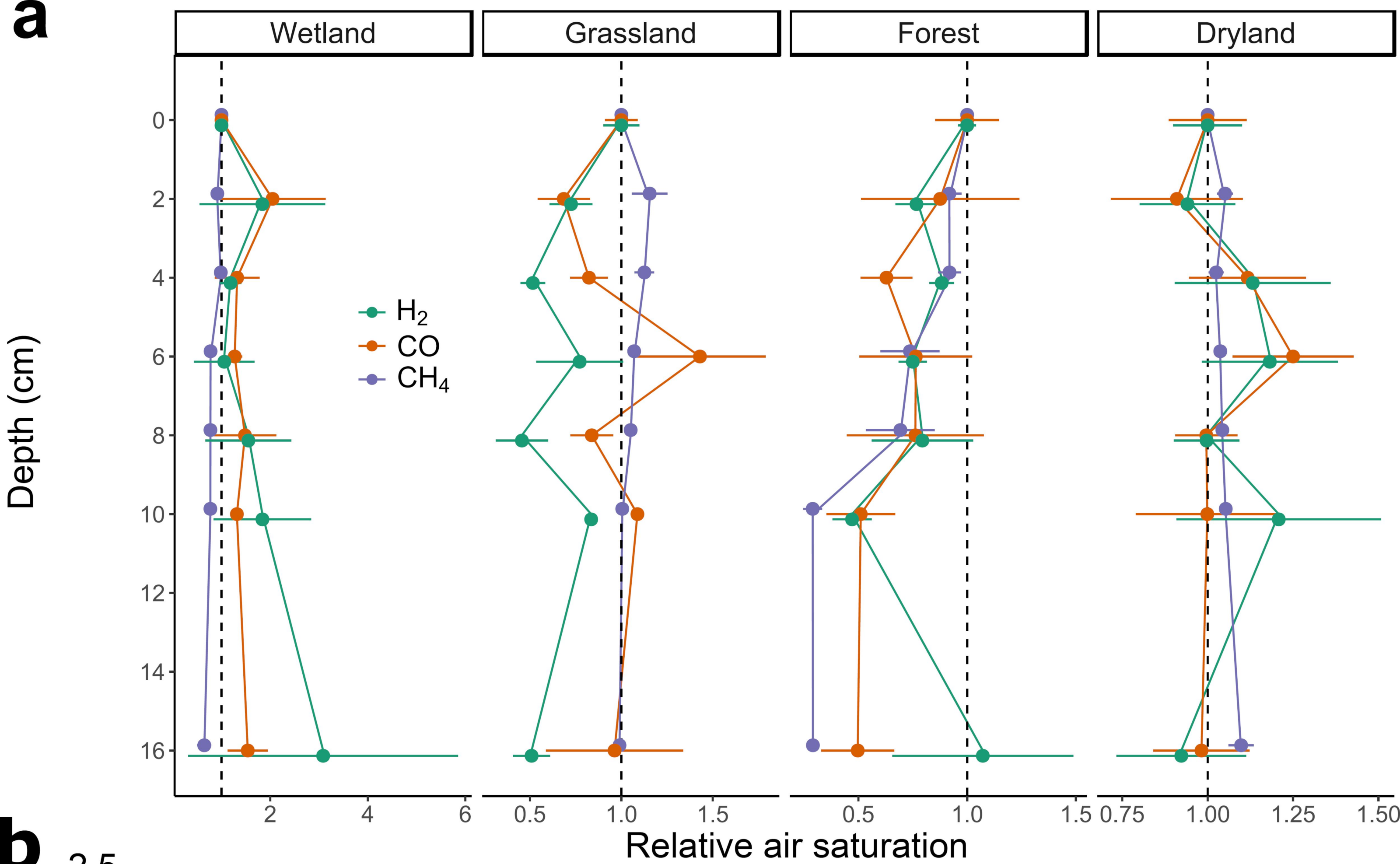




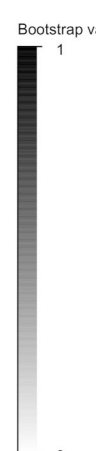
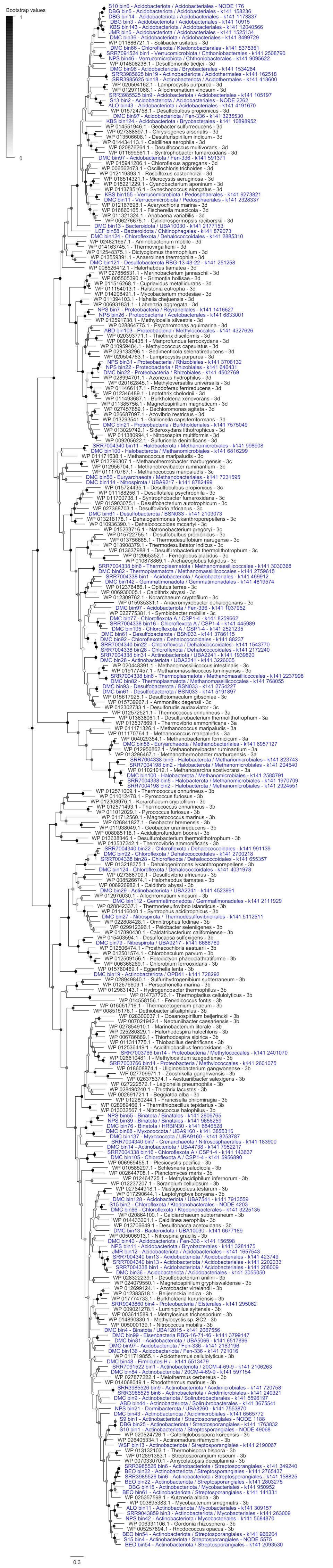




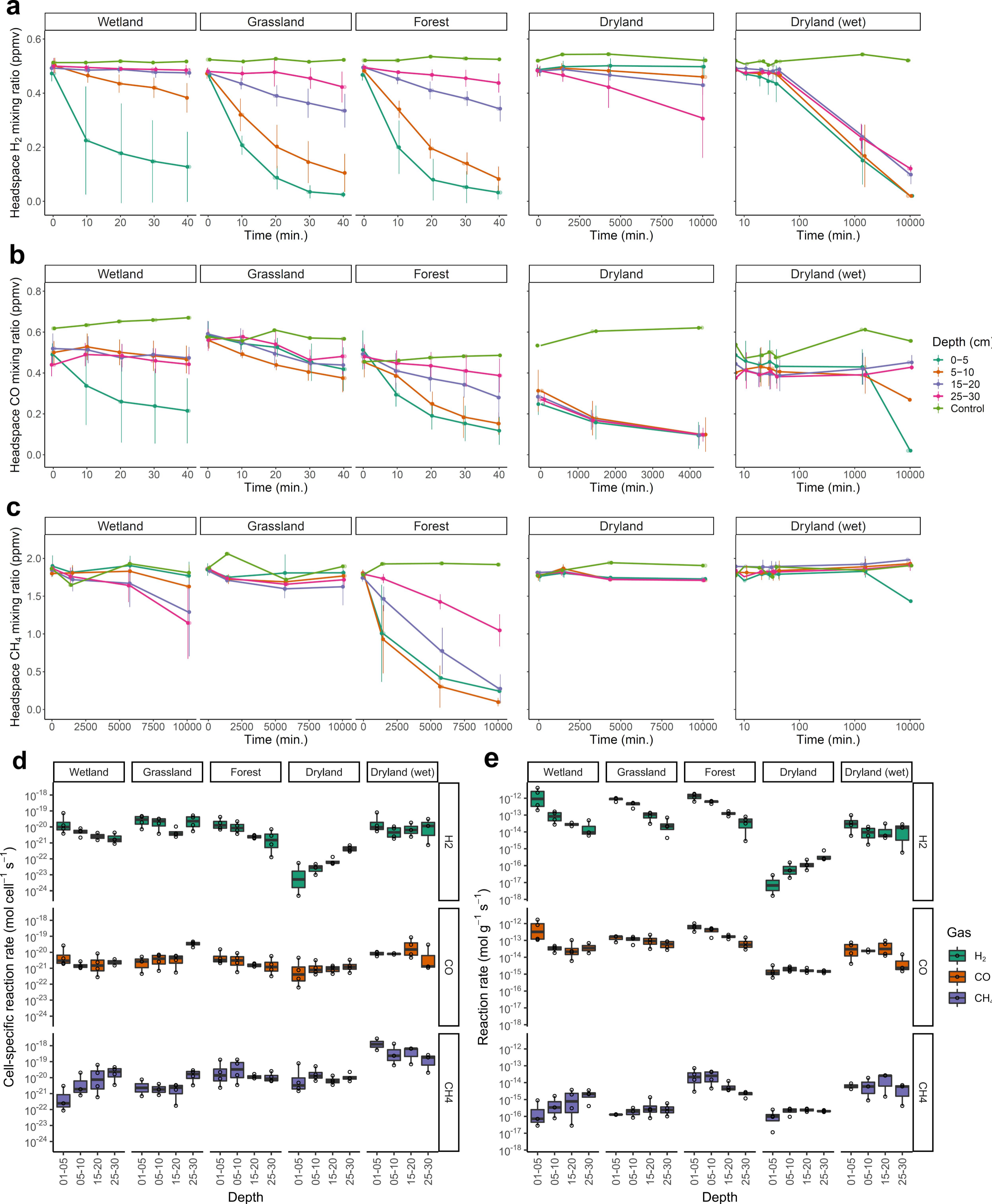




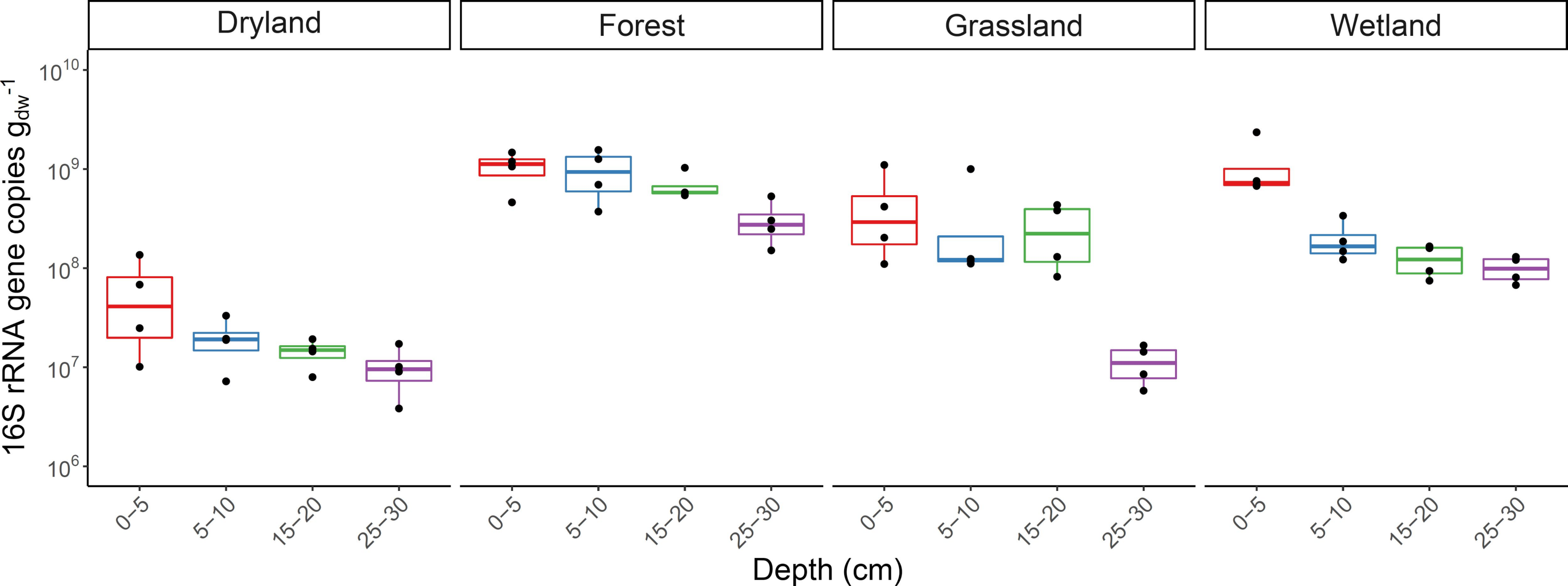


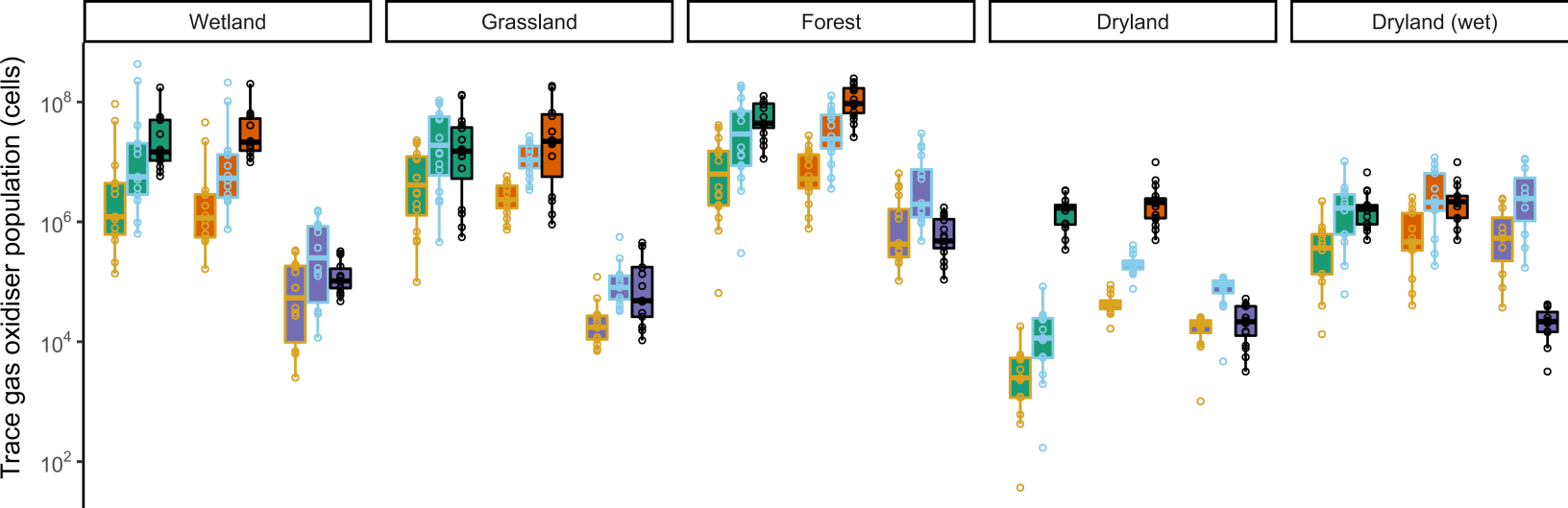




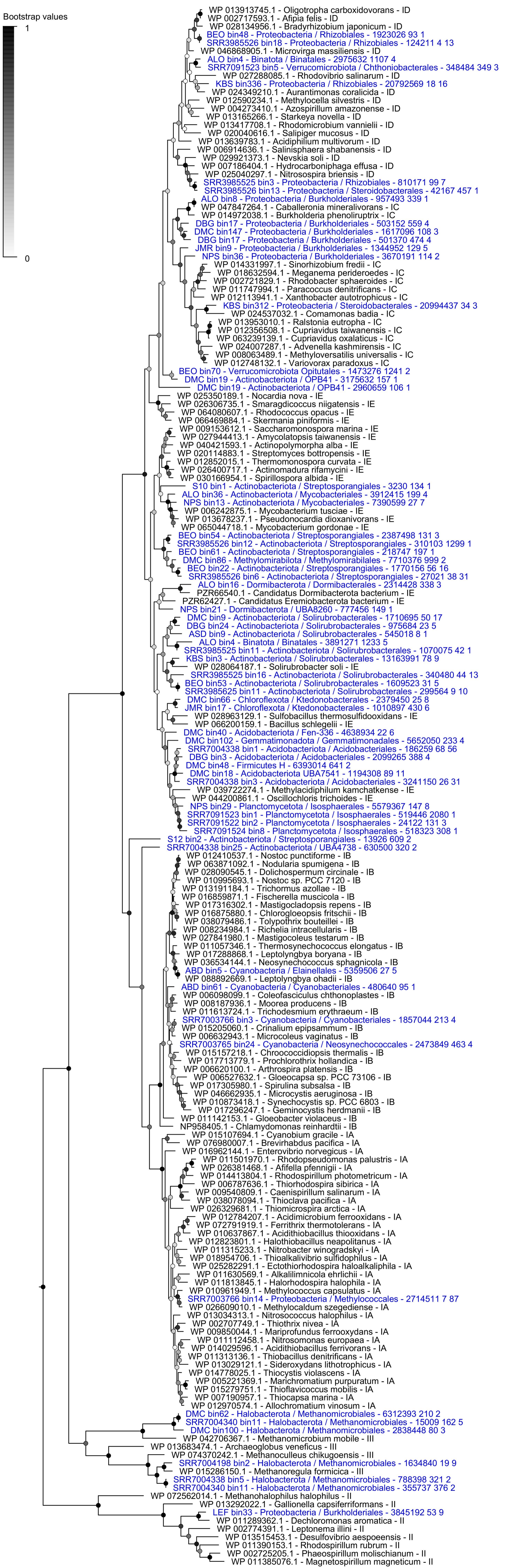






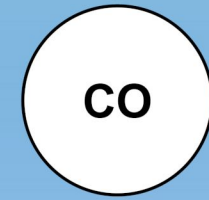
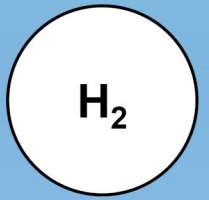








ATMOSPHERE



70  
Tg y<sup>-1</sup>

250  
Tg y<sup>-1</sup>

30  
Tg y<sup>-1</sup>

TRACE GAS OXIDIZERS

- Acidobacteriota
- Actinobacteriota
- Proteobacteria
- Gemmatimonadota
- Verrucomicrobiota
- Planctomycetota
- Myxococcota
- Nitrospirota
- Bacteroidota

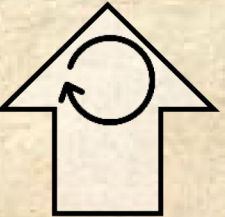
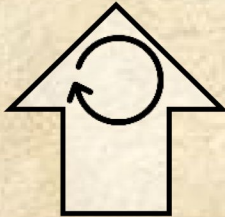
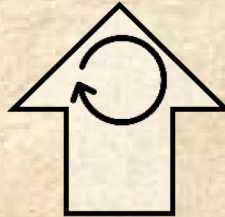
- Binatota
- Chloroflexota
- Desulfuromonadota
- Dormibacterota
- Eisenbacteria
- Eremiobacterota
- Methylomirabilota
- UBP7 A

- Actinobacteriota
- Proteobacteria
- Acidobacteriota
- Dormibacterota
- Eremiobacterota
- Methylomirabilota
- Myxococcota

- Chloroflexota
- Chloroflexota A
- Firmicutes H
- Gemmatimonadota
- Tectomicrobia
- Verrucomicrobiota

- Proteobacteria
- Gemmatimonadota  
(Methylomirabilota)
- (Verrucomicrobiota)

O<sub>2</sub>



NITROGEN FIXATION

THERMAL/PHOTOCHEMICAL DEGRADATION

METHANOGENESIS

GEOLOGICAL SOURCE  
(e.g., serpentinization)

FERMENTATION

GEOLOGICAL SOURCE  
(e.g., natural gas seeps)

SOIL