

# UC Davis

## UC Davis Previously Published Works

### Title

Diverse sediment microbiota shape methane emission temperature sensitivity in Arctic lakes

### Permalink

<https://escholarship.org/uc/item/3gd6x7mc>

### Journal

Nature Communications, 12(1)

### ISSN

2041-1723

### Authors

Emerson, Joanne B  
Varner, Ruth K  
Wik, Martin  
et al.

### Publication Date














2021

### DOI

10.1038/s41467-021-25983-9

Peer reviewed

# Diverse sediment microbiota shape methane emission temperature sensitivity in Arctic lakes

Joanne B. Emerson <sup>1,10</sup>, Ruth K. Varner <sup>2,3</sup>, Martin Wik<sup>4</sup>, Donovan H. Parks <sup>5</sup>,  
Rebecca B. Neumann <sup>6</sup>, Joel E. Johnson<sup>2</sup>, Caitlin M. Singleton <sup>5,11</sup>, Ben J. Woodcroft <sup>5</sup>,  
Rodney Tollerson II<sup>1,12</sup>, Akosua Owusu-Domney<sup>7,13</sup>, Morgan Binder<sup>7,14</sup>, Nancy L. Freitas <sup>7,15</sup>,  
Patrick M. Crill <sup>4</sup>, Scott R. Saleska<sup>8</sup>, Gene W. Tyson <sup>5,9</sup> & Virginia I. Rich <sup>1</sup>

Northern post-glacial lakes are significant, increasing sources of atmospheric carbon through ebullition (bubbling) of microbially-produced methane (CH<sub>4</sub>) from sediments. Ebullitive CH<sub>4</sub> flux correlates strongly with temperature, reflecting that solar radiation drives emissions. However, here we show that the slope of the temperature-CH<sub>4</sub> flux relationship differs spatially across two post-glacial lakes in Sweden. We compared these CH<sub>4</sub> emission patterns with sediment microbial (metagenomic and amplicon), isotopic, and geochemical data. The temperature-associated increase in CH<sub>4</sub> emissions was greater in lake middles—where methanogens were more abundant—than edges, and sediment communities were distinct between edges and middles. Microbial abundances, including those of CH<sub>4</sub>-cycling microorganisms and syntrophs, were predictive of porewater CH<sub>4</sub> concentrations. Results suggest that deeper lake regions, which currently emit less CH<sub>4</sub> than shallower edges, could add substantially to CH<sub>4</sub> emissions in a warmer Arctic and that CH<sub>4</sub> emission predictions may be improved by accounting for spatial variations in sediment microbiota.

<sup>1</sup> Department of Microbiology, The Ohio State University, 496W 12th Ave, Columbus, OH 43210, USA. <sup>2</sup> Department of Earth Sciences, University of New Hampshire, 56 College Road, Durham, NH 03824, USA. <sup>3</sup> Earth Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, 8 College Road, Durham, NH 03824, USA. <sup>4</sup> Department of Geological Sciences, Stockholm University, Stockholm 106 91, Sweden. <sup>5</sup> Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane 4072, Australia. <sup>6</sup> Civil & Environmental Engineering, University of Washington, 201 More Hall, Box 352700, Seattle, WA 98195-2700, USA. <sup>7</sup> Department of Environmental Science, University of Arizona, Tucson, AZ 85721, USA. <sup>8</sup> Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. <sup>9</sup> Centre for Microbiome Research, Queensland University of Technology, 37 Kent St, Woolloongabba, QLD 4102, Australia. <sup>10</sup> Present address: Department of Plant Pathology, University of California, Davis, One Shields Ave, Davis, CA 95616, USA. <sup>11</sup> Present address: Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg 9220, Denmark. <sup>12</sup> Present address: Department of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91106, USA. <sup>13</sup> Present address: Parkland Hospital, 5200 Harry Hines Blvd., Dallas, TX 75235, USA. <sup>14</sup> Present address: John C. Lincoln Health Network, 34975N North Valley Pkwy Ste 100, Phoenix, AZ 85086, USA. <sup>15</sup> Present address: Energy and Resources Group, University of California, Berkeley, USA. ✉email: [jbemerson@ucdavis.edu](mailto:jbemerson@ucdavis.edu); [ruth.varner@unh.edu](mailto:ruth.varner@unh.edu); [rich.270@osu.edu](mailto:rich.270@osu.edu)

At high latitudes, lakes and ponds are recognized as a large and understudied source of methane (CH<sub>4</sub>)<sup>1–4</sup>, a radiatively important trace gas. Post-glacial lakes (formed by glaciers and receding ice sheets, leaving mineral-rich sediments) represent the largest lake area at high latitudes<sup>5</sup>. Because of their areal extent, these lakes contribute to approximately two-thirds of the model-predicted natural CH<sub>4</sub> emissions above 50°N latitude<sup>1</sup>. Their geochemistry and emissions are distinct from thermokarst lakes formed by permafrost thaw<sup>6</sup>. With warming, permafrost thaw, and predicted increased precipitation, northern lakes are expected to receive more terrestrially derived carbon, likely increasing their carbon dioxide (CO<sub>2</sub>) and CH<sub>4</sub> emissions<sup>7,8</sup>.

Ebullition commonly accounts for >50%, sometimes >90%, of the CH<sub>4</sub> flux from post-glacial lakes, with the remainder primarily attributed to diffusion-limited hydrodynamic flux<sup>9,10</sup>. Ebullition moves CH<sub>4</sub> rapidly from sediments directly to the atmosphere, typically bypassing microbial CH<sub>4</sub> oxidation in the water column<sup>11</sup>. Incoming short-wave radiation and sediment temperature have been identified as strong predictors of ebullitive CH<sub>4</sub> emission from sub-arctic post-glacial lakes on an annual basis, with higher temperatures increasing emissions during the ice-free season<sup>2,12</sup>. However, the extent and drivers of spatial variability in this temperature response, particularly within lakes, are poorly understood.

To address this knowledge gap, we analyzed ebullitive CH<sub>4</sub> emissions over a 6-year period and collected underlying sediment cores in July 2012 from the littoral (edge) and pelagic (middle) locations of two shallow post-glacial lakes, Mellersta Harrsjön and Inre Harrsjön (Supplementary Fig. 1, Supplementary Table 1). These lakes are part of the Stordalen Mire complex, a hydrologically interconnected, discontinuous permafrost ecosystem encompassing post-glacial lakes and a mosaic *palsa*/wetland in approximately equal portions<sup>13</sup>. The lakes contribute ~55% of the total Stordalen Mire ecosystem CH<sub>4</sub> loss<sup>2</sup>, and are model sites for studying ebullitive CH<sub>4</sub> emissions via inverted funnel bubble traps at the lake surface<sup>9,12,14</sup>. These ebullitive flux measurements were collected for the six summers from 2009 to 2014<sup>12,14</sup> every 1–3 days<sup>9</sup>. Here, we analyzed 5126 ebullitive CH<sub>4</sub> emission measurements (Supplementary Table 2) from this previously published dataset for spatial patterns (edge vs. middle), and we linked these patterns to analyses of the microbiota and biogeochemistry in the underlying sediments.

## Results and discussion

### Spatial variation in ebullitive CH<sub>4</sub> temperature sensitivity.

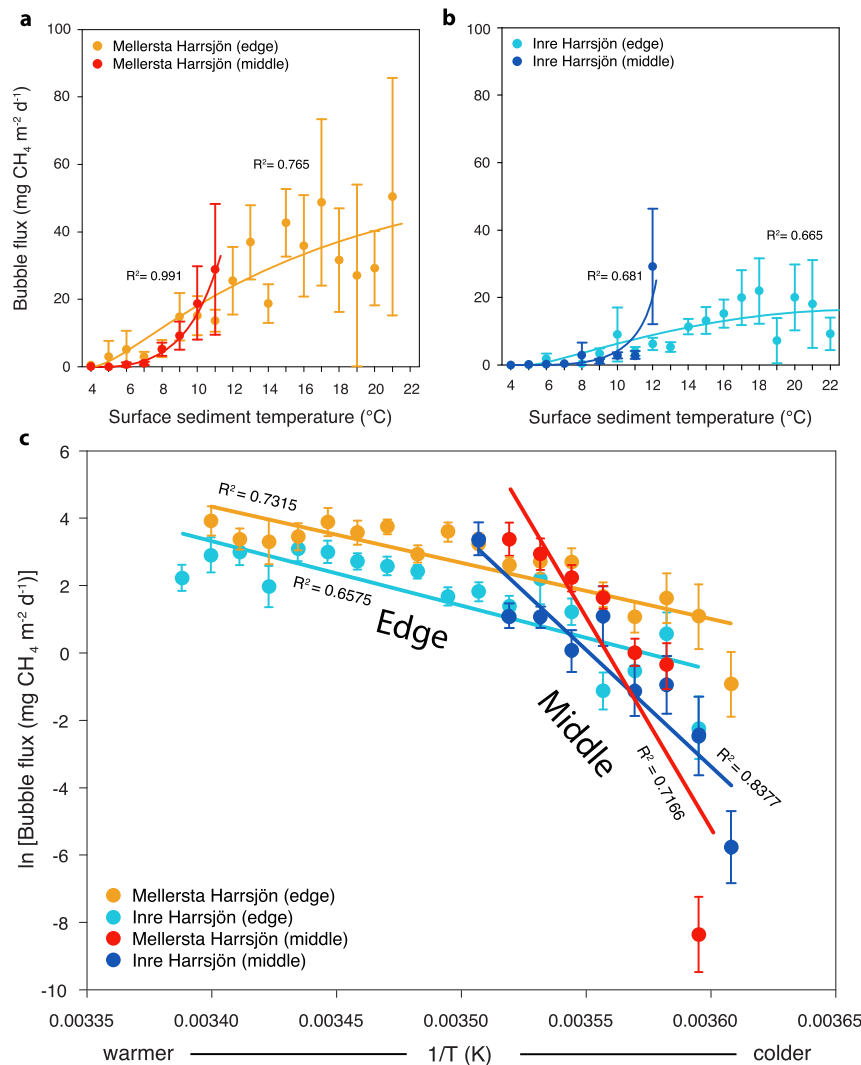
Previous work has shown that annual ebullitive emissions are consistently higher from these lakes' shallow littoral zones than their deeper pelagic zones<sup>9,15</sup>, as expected since the shallow sediments experience higher temperatures for longer periods and also receive more substrate input from aquatic vegetation<sup>16</sup>. However, assessing the temperature sensitivity of ebullition for the two lake zones in this study revealed a previously unnoticed significant difference, with ~3–5-fold higher temperature sensitivity in lake middles relative to edges (Fig. 1, Supplementary Table 3). The statistical significance of these differences was consistent across all edge-to-middle comparisons (within and between lakes) and was generally also robust to re-analyses considering three subsets of the data (removal of edge data at temperatures above those experienced by middles and removal of potential outliers at the highest and lowest temperatures experienced by lake middles, respectively), with one exception: Inre Harrsjön edges and middles were not statistically significantly different when higher temperature edge data were removed (Supplementary Table 3, Supplementary Fig. 2). Predicted future emissions from post-glacial subarctic lakes are based on current

measurements of temperature responsiveness<sup>1</sup>, which are dominated by ebullitive flux data from shallow lake zones because those locations currently experience a longer period of sufficient warmth for seasonal emissions than lake middles (~3 months relative to ~1 month)<sup>2</sup>. If, as suggested here by our spatially resolved emissions data, temperature responsiveness is substantially higher in the deeper sediments, then, as deeper regions warm and remain heated for longer before cooling off in the fall, future lake emissions would be greater than currently predicted. Thus, accurate CH<sub>4</sub> emission predictions rely on understanding the spatial heterogeneity and underlying causes of this temperature responsiveness.

Ebullition is controlled by CH<sub>4</sub> production (which is in turn driven by redox, substrates, temperature, and microbiota), consumption (driven by redox and microbiota)<sup>17–19</sup>, and the physics of bubble formation and escape (determined by sediment texture and overlying hydrostatic pressure, which is largely controlled by atmospheric conditions)<sup>2,15</sup>. Therefore, the edge-to-middle difference in temperature responsiveness of CH<sub>4</sub> ebullition could be partly due to differences in physicochemical characteristics (e.g., sediment texture, pressure, and redox), substrates (e.g., organic carbon), and/or microbiota (abundance, composition, and/or activity)<sup>20</sup>. Although differences in sediment texture were observed between the lake edge and middle in Mellersta Harrsjön, these differences were not consistent between lakes (Supplementary Fig. 3, Supplementary Table 4). Our previous work has shown higher and more variable ebullition rates during periods of dropping atmospheric pressure, but there were no differences in edge vs. middle locations<sup>9</sup>. In terms of redox, we expect concentrations of terminal electron acceptors to be low, as the likely source would be runoff<sup>21</sup>, and total sulfur and nitrogen did not correlate with ebullition rates by lake or location<sup>15</sup>. In terms of measured substrates, carbon:nitrogen (C:N) ratios and bulk δ<sup>13</sup>C<sub>TOC</sub> (indicative of vegetation composition) did not vary from edges to middles. Total organic carbon (TOC) varied by the lake, with similar concentrations observed between lake edge and middle in Mellersta and appreciably higher TOC in middle sediments in Inre Harrsjön. Carbon quality, as assessed by visual comparisons of organic matter composition, revealed coarse, less decomposed detritus *gyttja* (organic-rich, peat-derived mud) in the edge sediments of both lakes, while middle sediments were characterized by fine-grained, generally more decomposed detritus *gyttja*<sup>15</sup>. Thus, higher temperature responsiveness occurred where there was lower potential substrate quality, suggesting that substrate differences do not readily explain patterns in CH<sub>4</sub> emission responses to temperature in edge vs. middle lake locations, although more detailed substrate analyses could further evaluate this in the future.

**Spatial differences in sediment microbial communities.** Next, we sought to characterize differences in the microbiota that could contribute to the observed temperature response patterns in CH<sub>4</sub> emissions. We first used a 16S rRNA gene amplicon sequencing approach to characterize microbial community composition from the edge and middle cores from each lake (Fig. 2A, B, Supplementary Table 5). Although microbial community composition differed most significantly by depth within the sediment (Supplementary Fig. 4, Supplementary Table 6), as is typical for aquatic sediments<sup>22</sup>, significant differences between lake edges and middles (Fig. 2C, PERMANOVA *p* = 0.001) revealed shared spatial patterns in microbiota and ebullitive CH<sub>4</sub> flux measurements, suggesting that microbiota could contribute to the observed temperature sensitivity in CH<sub>4</sub> emissions.

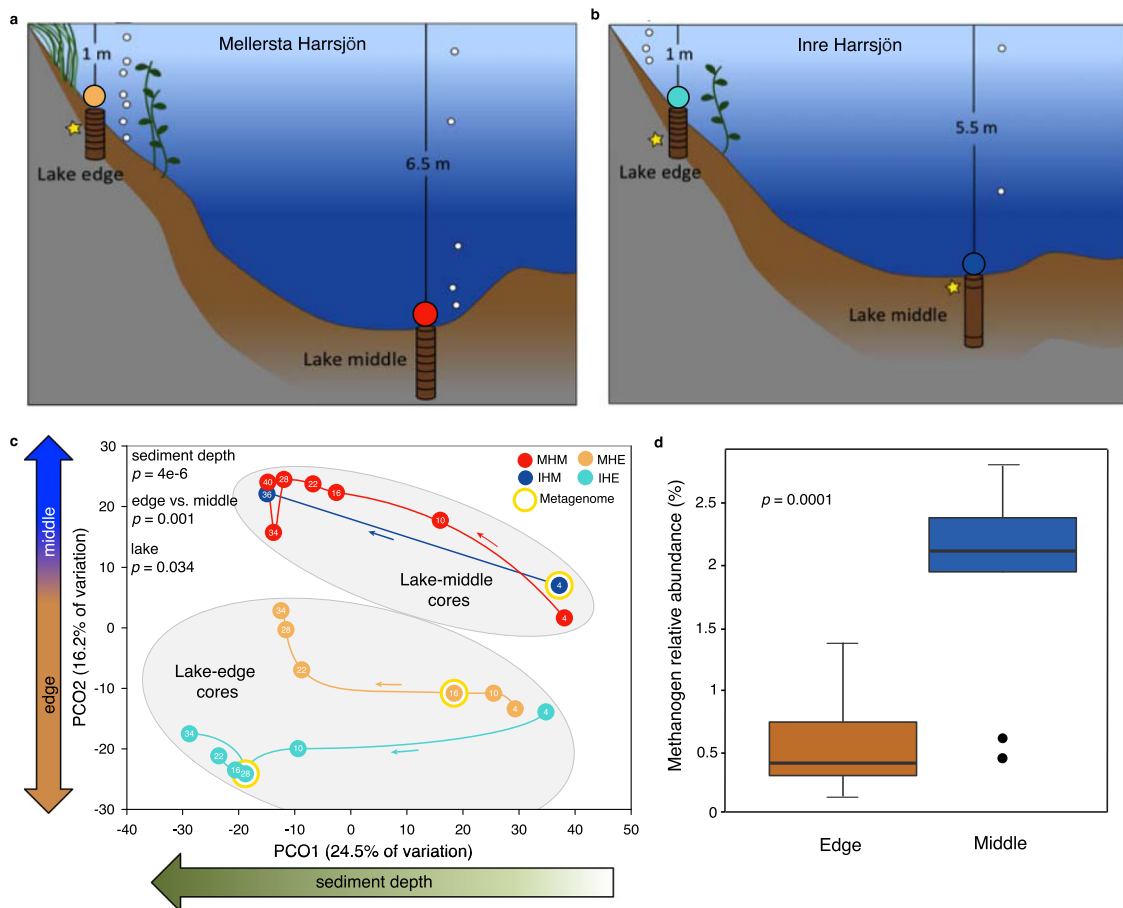
While total microbial abundances correlated most strongly with depth in the sediment and did not exhibit edge vs. middle



**Fig. 1 Temperature responsiveness of ebullitive methane flux from two post-glacial lakes.** Ebullitive CH<sub>4</sub> flux as a function of surface sediment temperature (data were binned in 1 °C intervals; see Methods) from June to September 2009–2014 for the edge vs. middle regions of **a** Lake Mellersta Harrsjön (MH) (MH edge—*n* = 1581, MH middle—*n* = 795 independent ebullitive CH<sub>4</sub> flux measurements) and **b** Lake Inre Harrsjön (IH) (IH edge—*n* = 2318, IH middle—*n* = 432 independent ebullitive CH<sub>4</sub> flux measurements). Error bars in **a** and **b** are 95% confidence intervals, fit lines are 2nd-degree polynomials, and points are means. **c** Arrhenius plots of the data in **a** and **b** (*n* = 5126 independent ebullitive CH<sub>4</sub> flux measurements); ln(bubble CH<sub>4</sub> flux) vs. the inverse surface sediment temperature in K. Points are means and error bars are 95% confidence intervals. Data are color-coded by the lake and by edge (littoral) and middle (pelagic) zones.

differences ( $p = 0.15$ , Supplementary Fig. 5, Supplementary Table 7), an exploration of functional guilds revealed multiple significant edges vs. middle differences. Methanogens (defined here as populations from known methanogenic clades<sup>23</sup>, Supplementary Table 5) were significantly more relatively abundant in lake middles than edges (Fig. 2D, ANOVA  $p = 0.0001$ ). Combining relative abundances and quantitative polymerase chain reaction (qPCR) absolute abundances as a proxy for methanogen total abundances showed no significant difference between lake edges and middles ( $p = 0.94$ ), likely due to the strong sediment-depth patterns of total cell numbers (qPCR abundances) in combination with the low relative abundances of methanogens (0.3–2.3%, consistent with other studies<sup>24–27</sup>). Syntrophs have been shown to be upstream regulators of methane production<sup>28–31</sup>, and they can be obligately mutualistic with methanogens, for example by producing hydrogen used in hydrogenotrophic methanogenesis. While syntrophy does not exhibit a strong phylogenetic signal<sup>32</sup>,

precluding its robust quantification here as a functional guild, the Syntrophaceae lineage, whose syntrophic potential was supported by reconstruction of a metagenome-assembled genome (MAG) (see below), did show significantly higher relative abundances in lake middle sediments relative to edges ( $p = 0.047$ ). Aerobic methanotrophs, which are posited to have minimal impact on ebullitive CH<sub>4</sub> loss due to rapid bubble movement through sediment<sup>11</sup>, were confined to the surface sediment layers as expected (Supplementary Table 8) and did not differ significantly in composition or relative abundance between edges and middles (ANOVA  $p = 0.76$ ). Anaerobic methanotroph (ANME archaeal) abundances differed significantly between lake edges and middles (ANOVA  $p = 0.014$ ) and were approximately one order of magnitude higher in edge sediments (Supplementary Tables 8 and 9). Although this could suggest that increased anaerobic methane oxidation in the edge sediments could contribute to the observed differences in temperature sensitivity, these ANME archaea comprised only 0.1% of the community on



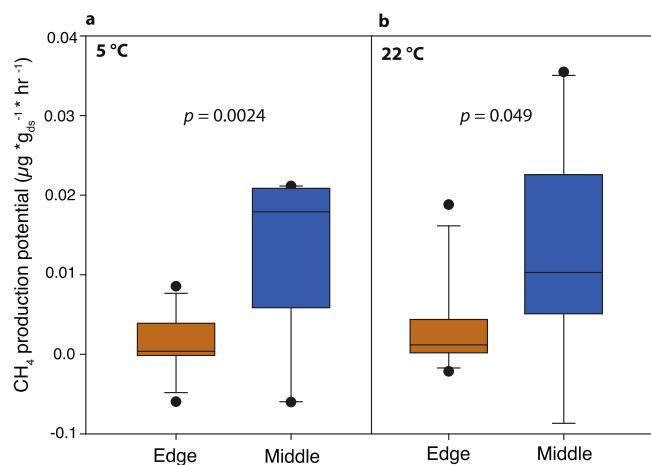
**Fig. 2** Lake sediment bacteria and archaea in two post-glacial lakes. **a, b** Schematic overview of lakes and cores collected for DNA sequencing analyses, with core subsections indicated by horizontal lines. Cores in each lake are referred to as Lake edge or Lake middle, with overlying water depth as indicated, and the four colored circles are used to distinguish each core and/or lake location throughout the figures. Yellow stars indicate cores and depths targeted for shotgun metagenomics. **c** Principal coordinates analysis (PCoA) of microbial community composition across samples (each core subsection,  $n = 21$ ), based on 16S rRNA gene amplicon abundances of microbial operational taxonomic units (OTUs); circles represent samples, and samples in closer proximity have more similar microbial community composition. Thin arrows along colored lines indicate increasing depth within each core.  $p$ -Values from one-way PERMANOVA indicate how significantly microbial community composition differed according to the indicated categorical variable (significant if  $p < 0.05$ ). **d** Percent relative abundance of OTUs identified as methanogens in 16S rRNA gene amplicon data in lake edges compared to lake middles ( $p$ -value from two-tailed Student's  $t$ -test, significant if  $p < 0.05$ ),  $n = 21$  biologically independent samples. Lines in boxes depict the median, boxes indicate 75th percentile, whiskers 95th percentile, and points are outliers.

average (range 0.0–0.6%, Supplementary Tables 5 and 8), and again, ebullition is expected to largely bypass methane oxidation. Still, future work to further constrain ANME activity at different temperatures<sup>33,34</sup> may help to elucidate the significance of the observed patterns in ANME abundance.

**Sediment incubations and modeling of CH<sub>4</sub> production.** To more directly evaluate the CH<sub>4</sub> production potential of the microbial communities in these sediments at different temperatures, we performed 48 ex situ anaerobic incubations of edge and middle sediments collected in 2012 (linked directly to our microbial and biogeochemical data) and 2013 (from the same four core locations) (Supplementary Table 10). These incubations at 5 and 22 °C showed that the lake-middle sediments had significantly higher CH<sub>4</sub> production potentials than lake-edge sediments at both temperatures (Fig. 3), paralleling their higher methanogen and syntroph relative abundances and indicating that the lake-middle methanogens can remain metabolically active at higher temperatures, despite never yet experiencing them in situ. However, it is important to note that a difference in temperature responsiveness (i.e., different CH<sub>4</sub> production

potentials at 5 °C, relative to 22 °C) was not observed for the edge or middle sediments (ANOVA  $p = 0.38$  for edges,  $p = 0.91$  for middles), as might have been expected from the in situ ebullition data (Fig. 1). One potential explanation for this is that the ex situ incubations measured relatively immediate CH<sub>4</sub> production when the entire sediment was maintained at a certain temperature, whereas the in situ measurements linked temperatures at the sediment/surface interface with ebullitive CH<sub>4</sub> flux measurements at the lake surface. Thus, the measured in situ temperatures were somewhat disconnected from the temperatures (and times) of CH<sub>4</sub> production, and their associated flux measurements represent CH<sub>4</sub> production potential integrated over the time scales of bubble generation and flow through the sediment and water column, as observed over multiple years.

Since lake surface ebullition measurements represent CH<sub>4</sub> production integrated across depths in the underlying sediment, we applied isotope and mass balance calculations to partition CH<sub>4</sub> production to its likely source depths, in order to link CH<sub>4</sub> production to depth-discrete microbiota. Based on stable carbon isotope values and porewater concentrations of CH<sub>4</sub> and dissolved inorganic carbon (DIC), we inferred total CH<sub>4</sub> loss (fugitive CH<sub>4</sub>) at each depth interval (Supplementary Table 4).



**Fig. 3 Methane production from anaerobic laboratory incubations of lake sediments.** Sediments were collected from edges and middles of lakes Inre Harrsjön and Mellersta Harrsjön in 2012 and 2013 (see Methods) and incubated at **a** 5 °C ( $n = 12$  independent incubation experiments) and **b** 22 °C ( $n = 12$  independent incubation experiments). Headspace  $\text{CH}_4$  concentrations were measured daily for 5 days and average daily  $\text{CH}_4$  fluxes were calculated for each sample. Lines in boxes depict the median, boxes indicate 75th percentile, whiskers 95th percentile, and points are outliers. ds dry sediment.  $p$ -Values for both **a** and **b** are from one-way ANOVA edge vs. middle comparisons (significant if  $p < 0.05$ ).

This fugitive  $\text{CH}_4$  production was correlated with microbiota from the same depths, revealing a significant correlation between overall microbial community composition and fugitive  $\text{CH}_4$  (Mantel  $p = 0.016$ ) and between fugitive  $\text{CH}_4$  and the relative abundances of specific microbial lineages, including Phycisphaerae, Thermoplasmata, and Aminicenantes (Supplementary Tables 6 and 11). All of these lineages have representative MAGs with metabolic reconstructions discussed further below.

### Microbial metabolic predictions and ties to biogeochemistry.

To more specifically investigate links between  $\text{CH}_4$ -associated microbial functional guilds and  $\text{CH}_4$  chemistry, we identified multiple known  $\text{CH}_4$ -cycling clades in the 16S rRNA gene amplicon data and applied targeted metagenomic sequencing to a subset of samples to examine diagnostic genes for  $\text{CH}_4$  cycling (and to assemble genomes for metabolic pathway reconstructions, discussed further below). From the metagenomes, using hidden Markov models (see Methods), we recovered 5470 sequencing reads with high similarity to phylogenetically diverse functional genes indicative of  $\text{CH}_4$  production (*mcrA*) and consumption (*pmoA*) potential (Supplementary Fig. 6, Supplementary Table 12). The *mcrA* data, together with 16S rRNA gene abundances of specific lineages of known hydrogenotrophic and acetoclastic methanogens, support  $\alpha\text{C}$  isotopic calculations in indicating that hydrogenotrophic methanogenesis was the dominant methanogenic pathway in these lake sediments (Supplementary Fig. 7).

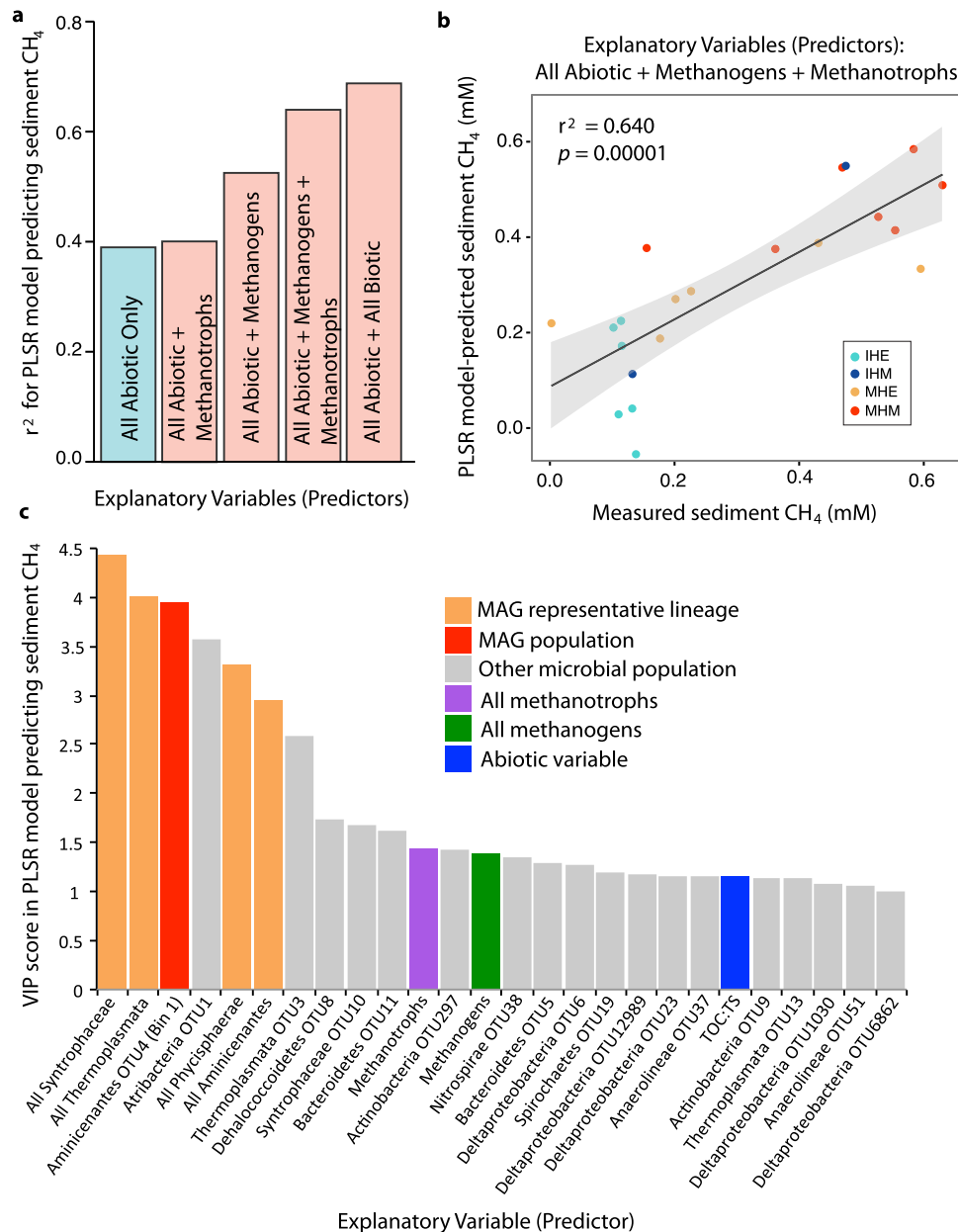
Using partial least squares regressions (PLSR) and multiple linear regression (MLR) analyses, we predicted porewater  $\text{CH}_4$  concentrations from methanogen and methanotroph relative abundances, as measured via 16S rRNA gene amplicon sequencing data. Although the modeled fugitive  $\text{CH}_4$  calculations might have provided a better depth-resolved link to  $\text{CH}_4$  production than the porewater  $\text{CH}_4$  concentrations analyzed here, we opted to use two direct measurements (microbiota and porewater  $\text{CH}_4$ ) in these statistical models, rather than essentially modeling a model. When using either PLSR or MLR to predict porewater

$\text{CH}_4$  concentrations, a better prediction was achieved when both depth-resolved abiotic variables (i.e., depth, TOC, DIC,  $^{13}\text{C}_{\text{TOC}}$ , S, and TOC:TS, see Methods) and the relative abundances of predicted  $\text{CH}_4$ -cycling organisms were included (PLSR:  $r^2 = 0.640$ ,  $p = 0.00001$ , MLR: adjusted  $r^2 = 0.752$ ,  $p = 0.0003$ ), relative to including the abiotic variables alone (PLSR:  $r^2 = 0.390$ ,  $p = 0.002$ , MLR: adjusted  $r^2 = 0.532$ ,  $p = 0.0004$ ) (Fig. 4A, B, Supplementary Table 13). These results suggest that direct measurements of microbial abundances could contribute to more accurate predictions of future  $\text{CH}_4$  emissions, consistent with previous statistical models that have linked specific microbiota to C- and/or  $\text{CH}_4$ -cycling dynamics in marine ecosystems and thawing permafrost peatlands<sup>35–39</sup>.

By expanding our PLSR analyses to consider the full microbial community, in addition to known  $\text{CH}_4$  cycles, our ability to predict  $\text{CH}_4$  concentrations improved further. This analysis considered the following groupings of 16S rRNA gene abundances as explanatory variables for the prediction of porewater  $\text{CH}_4$  concentrations: (1) each operational taxonomic unit (OTU) at  $>1\%$  relative abundance in any sample (Supplementary Table 5), (2) summed lineage abundances of all bacteria and archaea (mostly at the phylum or class levels, see Supplementary Fig. 4 for groupings), and (3) summed abundances of the most highly resolved lineage representative in the amplicon data for each MAG (a population genome computationally reconstructed from shotgun metagenomic community DNA sequencing data, Supplementary Table 14). In two cases (for Aminicenantes MAG Bin 1/OTU 4 and a Methanomassiliococcales-related Thermoplasmata, MAG Bin 16/OTU 27), a MAG was linked directly to a specific OTU in the amplicon data through a co-binned 16S rRNA gene sequence in the MAG, such that the MAG relative abundance could be inferred from the amplicon data. In all other cases, the summed abundances of amplicon OTUs in the same lineage as the MAG were used as proxies for MAG abundances. A total of 153 potential explanatory variables were considered in this PLSR analysis, 26 of which, including methanogen and methanotroph abundances, were identified as significant predictors of porewater  $\text{CH}_4$  concentrations (Fig. 4C).

Four of the top five microbial groups most predictive of porewater  $\text{CH}_4$  concentrations in the PLSR analysis were lineages for which we were able to reconstruct a MAG (Fig. 4C, Supplementary Tables 15 and 16), thus organization into MAGs helped to unravel the specific metabolic processes most predictive of carbon chemistry. In total, five MAGs were reconstructed with  $>85\%$  completeness and  $<6\%$  contamination (Supplementary Discussion). The best overall predictor of porewater  $\text{CH}_4$  concentrations was the Syntrophaceae class of Deltaproteobacteria, which was considered in the PLSR analysis as the summed abundance of all OTUs in this clade. Consistent with a syntrophic metabolism, including hydrogen production (e.g., in support of hydrogenotrophic methanogenesis, which is the dominant methanogenic pathway in these sediments, as described above), the Syntrophaceae MAG revealed 15 hydrogenase-associated genes, along with the capacity to ferment diverse carbon compounds (particularly carbon-sulfur compounds), with the added potential capacity for respiration (see Supplementary Discussion).

Though the Syntrophaceae lineage was overall most predictive of porewater  $\text{CH}_4$  concentrations, the most significant predictive single OTU was a member of the candidate phylum Aminicenantes, which we also recovered as a MAG. While this lineage has been previously predicted to be fermentative, saccharolytic, and/or aerobic<sup>40–42</sup>, our lake sediment genome revealed the metabolic potential for several C1 metabolic processes, including methylotrophy through the assimilation of methylamines, methane-thiols, and/or dimethylsulfide, similar to previous recoveries of complete



**Fig. 4 Partial least squares regression (PLSR) statistical modeling to predict sediment  $\text{CH}_4$  concentrations.** PLSR analyses tested the ability of different suites of explanatory variables to predict measured sediment  $\text{CH}_4$  concentrations in the four cores from 2012 across depths ( $n = 21$ ); in all models, all measured abiotic variables (except those related to  $\text{CH}_4$  concentrations, see Methods) were included as explanatory variables, and biotic variables were added as indicated. Biotic variables included relative abundances of specific OTUs and/or summed OTU abundances grouped by taxonomy or predicted metabolism (as indicated), from 16S rRNA gene amplicon data. **a** Correlation coefficient ( $r^2$ ) for PLSR models predicting sediment  $\text{CH}_4$  using different combinations of explanatory variables. Each bar represents a single underlying data point, with the value of that point indicated by the bar height along the y-axis. **b** Linear regression of measured and model-predicted sediment  $\text{CH}_4$ , considering all abiotic variables and methanogen and methanotroph abundances as explanatory variables; error band represents 95% confidence interval; each point is a sample, colored by core. **c** For the model with the highest  $r^2$  (rightmost in panel **a**), all significant explanatory variables are shown (VIP scores  $> 1$ ,  $n = 26$  significant explanatory variables out of 153 total variables considered). VIP scores show the relative contribution of each variable to the model, with higher VIP scores indicating a more significant contribution. Each bar represents a single underlying data point, with the value of that point indicated by the bar height along the y-axis.

Wood-Ljungdahl pathways for C1 metabolism via carbonyl and methyl pathways in this lineage<sup>43</sup>. The predicted capacity for methylotrophy could explain the strong correlation between Aminicenantes relative abundance and porewater  $\text{CH}_4$  concentrations.

The relative abundances of two other lineages with MAGs, the Thermoplasmata (a group of Archaea) and Phycisphaerae (a class of Planctomycetes bacteria), were also strongly predictive of both

porewater  $\text{CH}_4$  concentrations in the PLSR analysis and of calculated fugitive  $\text{CH}_4$  in linear regressions (Supplementary Tables 11 and 16). Phylogenetic analyses showed that the Thermoplasmata Bin 19 MAG was derived from a divergent member of the Thermoplasmatales order, and it encodes the capacity for  $\text{CO}_2$  production from a formate, along with peptide and amino acid degradation (as previously indicated<sup>44</sup>) and complex carbon degradation. Our recovered Phycisphaerae

population genome appears to have the capacity to metabolize a wide variety of complex carbon compounds, potentially via fermentation, consistent with previous predictions for the Planctomycetes phylum<sup>45</sup>. While direct ties to CH<sub>4</sub> are not obvious in these two genomes, we speculate that their contributions to overall carbon cycling, potentially including through fermentative contributions of CO<sub>2</sub> as a substrate, may be driving these strong correlations with both CH<sub>4</sub> concentrations and modeled fugitive CH<sub>4</sub>.

Interestingly, the only lineage represented by a MAG that was not a significant predictor of porewater CH<sub>4</sub> concentrations in the PLSR analysis was classified by its 16S rRNA gene sequence as a member of the Methanomassiliococcales, an archaeal lineage presumed to consist exclusively of obligate H<sub>2</sub>-dependent methylophilic methanogens<sup>46,47</sup>. We hypothesized that this lake sediment population did not have the capacity for methanogenesis, as we did not recover genes from the methanogenesis pathway in this 95% complete genome. Further analysis of this MAG, reported in a separate paper<sup>48</sup>, revealed that this population is not a member of the Methanomassiliococcales but rather is part of a new order basal to the Methanomassiliococcales within the Thermoplasmata. All 12 MAGs that we analyzed from this new order and related lineages (including 11 MAGs from other ecosystems, mostly anoxic sediments) were found to lack >20 methanogenesis biomarker genes that are present in Methanomassiliococcales genomes, including *mcrA*. Instead, the lake sediment MAG reported here is predicted to conserve energy through amino acid metabolism<sup>48</sup>. While *mcrA* sequences that were putatively phylogenetically assigned to Methanomassiliococcales were recovered from metagenomic reads in our GraftM analysis here (Supplementary Table 12), these genes were presumably derived from other OTUs classified as Methanomassiliococcales in our amplicon data, some of which may be true methanogens. These results highlight the value of reconstructing near-complete genomes for a better understanding of metabolism, along with the limitations of using 16S rRNA gene sequences to infer metabolic processes.

In conclusion, we found significant differences in the slope of the temperature vs. CH<sub>4</sub> flux relationship between sub-arctic lake edges and middles, suggesting that radiative forcing (temperature) and a concomitant increase in microbial metabolic rates are not the only primary controls on CH<sub>4</sub> emissions. These edge vs. middle differences were shared by underlying sediment microbial communities, suggesting that differences in sediment microbial community composition (i.e., not simply differences in microbial activity) contribute to spatial differences in the response of CH<sub>4</sub> emissions to increasing temperature. Specifically, we observed significant differences in microbial community composition between lake edges and middles, including significantly higher methanogen and Syntrophaceae abundances in lake middles relative to edges, and CH<sub>4</sub> emissions from lake middle sediments were significantly higher than from lake edges when incubated at the same temperatures. In addition, the relative abundances of CH<sub>4</sub>-cycling organisms and their reconstructed population genomes (MAGs) were significantly better predictors of sediment CH<sub>4</sub> concentrations than abiotic variables alone. Syntrophic lineages, which can generate the hydrogen required for hydrogenotrophic methanogenesis, and lineages capable of C degradation to CO<sub>2</sub> (also potentially “upstream” of methanogenesis) were also predictive of sediment CH<sub>4</sub> concentrations. Together, these results suggest that when lake middles reach the temperatures of lake edges, they may emit even more CH<sub>4</sub> than the lake edges currently do, such that our projected future CH<sub>4</sub> emissions may be underestimating contributions from subarctic lakes, and that knowledge of microbial community composition and metabolism could improve these predictions. Future investigations that

consider the combined effects of microbiota, carbon quality, and temperature on lake CH<sub>4</sub> emissions, including further exploration of sediment depth-resolved contributions to total CH<sub>4</sub> emissions across multiple locations and years, will help to provide a more comprehensive understanding of spatiotemporal controls on global CH<sub>4</sub> emissions.

## Methods

**Field site and sample collection.** Stordalen Mire is a subarctic peatland complex located 10 km east of Abisko in northern Sweden (68°21'N, 19°02'E). Lakes Mellersta Harrsjön and Inre Harrsjön are 1.1 and 2.3 ha in area, reaching maximum depths of 7 and 5 m, respectively<sup>49</sup>. These lakes are post-glacially formed, and, unlike some thermokarst lakes at high latitude, they do not have underlying permafrost or seeps that emit geologic CH<sub>4</sub>; the CH<sub>4</sub> is biologically derived<sup>9,21</sup>. Mellersta Harrsjön receives water from a small stream, while Inre Harrsjön is fed through groundwater and runoff from the surrounding mire. Ebullitive and diffusion-limited CH<sub>4</sub> emissions from these lakes have been documented using floating funnels and chambers distributed across the lakes, and sampled frequently<sup>2,9,12</sup>. Ebullition varies spatially with higher emissions from shallow zones and in the presence of plants<sup>9,15</sup>.

We collected quadruplicate sediment cores (four cores from two locations in each of two lakes: Mellersta Harrsjön edge (68°35'7832"N, 19°04'2046"E) and middle (68°35'8291"N, 19°04'2132"E) and Inre Harrsjön edge (68°35'7880"N, 19°04'8525"E) and middle (68°35'8418"N, 19°04'5650"E)) on July 10 and 18, 2012 at the Stordalen Mire nature reserve, a research site near Abisko, northern Sweden (Supplementary Table 1). Samples were taken from cores (as described below) along a depth gradient (ranging from 4 to 40 cm) for geochemical measurements and microbial DNA sequencing data.

**Geochemical data collection and analysis.** For each set of four cores, we sampled the first core for sediment C, N, and S (weight percent), percent TOC, and bulk sediment <sup>13</sup>C<sub>TOC</sub> and <sup>15</sup>N<sub>TOC</sub>. Samples of 1 cm<sup>3</sup> were taken in 6 cm increments from the top of the core to the bottom. The samples were then dried, ground, and split into an untreated sample for total carbon (C) and an acidified TOC sample. Details regarding sample preparation for measurement on a Perkin Elmer 2400 Series II CHNS/O Elemental Analyzer at the University of New Hampshire (UNH) were described previously<sup>15</sup>. Repeatability error was established by analyzing replicate samples and calculating the standard deviation. Duplicate samples were run approximately every 10 samples. Potential outliers were also run in duplicate. Isotopic analysis was performed by combusting dried sediment samples in a Costech ECS 4010 elemental analyzer coupled to a Thermo Trace GC Ultra isotope ratio mass spectrometer (IRMS), based on calibration with acetanilide, Atlantic cod, black spruce needles, sorghum flour, corn gluten, NIST 1515 apple leaves, and tuna muscle standards (UNH Stable Isotope Lab). In 2013 we also collected sediment cores in the same locations in these lakes. We report sediment textural analyses from these cores as % sand, % silt, and % clay (Supplementary Table 4). Those samples were dried and run through a laser particle size analyzer (Malvern Mastersizer 2000).

The second replicate core was used for quantifying total CH<sub>4</sub> in the core sediment reported in μM. After coring, we pulled 2 cm<sup>3</sup> sediment plugs using cut plastic syringes through pre-drilled holes cut at 4 cm increments along the core liner. The sediment plugs were transferred to 30 ml serum vials containing 5 ml of 2 M NaOH, capped quickly, and shaken<sup>50,51</sup>. After sitting overnight then heating for 1 h at 60 °C, the headspace of the vials was analyzed for CH<sub>4</sub> using a Shimadzu GC-2014 gas chromatograph (GC) with a flame ionizing detector<sup>9</sup>. The CH<sub>4</sub> measured represents the total, that is, nearly all of the CH<sub>4</sub> dissolved in the water from the sediment plug and any bubbles that may have been trapped in the sediment. The remaining sediment samples in the vials were weighed and dried to constant weight to determine the mass of water in the samples to be used for calculating the CH<sub>4</sub> concentration in μM.

The third replicate core was used for the measurement of DIC. Rhizon samplers were inserted every 2 cm through pre-drilled holes in the core and a vacuum was pulled with a 30 ml polypropylene syringe. The first ~1 ml of sediment water was discarded because of contamination with DI water. After 10 ml of sediment pore water was collected, it was injected into a 30 ml evacuated serum vial with 1 ml 30% H<sub>3</sub>PO<sub>4</sub> solution. This caused forms of inorganic C in the water to form CO<sub>2</sub>. A headspace sample was then extracted and run on an infrared gas analyzer to determine the CO<sub>2</sub> concentration.

Porewater isotopic composition was determined in samples from cores collected in the same locations in 2014. Methods were described previously<sup>35</sup>. Briefly, sample vials that were collected for CH<sub>4</sub> and DIC were acidified with 0.5 ml of 21% H<sub>3</sub>PO<sub>4</sub> and brought to atmospheric pressure with helium. The sample headspace was analyzed for δ<sup>13</sup>C of CH<sub>4</sub> and CO<sub>2</sub> on a continuous-flow Hewlett-Packard 5890 gas chromatograph (Agilent Technologies) at 40 °C coupled to a Finnigan MAT Delta S isotope ratio mass spectrometer via a ConFlo IV interface system (Thermo Scientific).

Methods for measuring ebullition and water temperature have been described previously, and the ebullitive CH<sub>4</sub> flux and temperature data analyzed here are a subset of those previously reported<sup>9</sup>. In brief, measurements of CH<sub>4</sub> bubble flux



during the ice-free season (June to September) have been ongoing at these lakes since 2009. A total of 40 bubble traps (here, 22 traps), distributed in a depth-stratified sampling scheme were sampled frequently (every 1–3 days). For this study, averages of CH<sub>4</sub> bubble flux were calculated for each lake by binning data from the edge (shallower water, littoral) and middle (deeper water, pelagic) areas separately in 1 °C interval (total of 4–22 °C) of corresponding surface sediment temperature. For this, we used flux and temperature data collected from 2009–2014. Water and surface sediment temperatures were measured in profiles continuously, using intercalibrated Onset HOBO v22 loggers, as previously described<sup>9</sup> (data are available here: <https://bolin.su.se/data/>). The binned flux data were used to construct Arrhenius equations in order to investigate differences in temperature response on the ebullition from the edge and middle areas.

**DNA extraction and 16S rRNA gene sequencing.** A fourth replicate core was collected for DNA extraction. After coring, we pulled 2 cm<sup>3</sup> sediment plugs using cut plastic syringes through pre-drilled holes cut at 4 cm increments along the core liner. Samples were immediately put in Eppendorf tubes and placed in a cooler until returned to the research station where they were stored at –80 °C until extraction.

For DNA extraction from each core depth range, 0.25 g of sediment was collected under sterile conditions and added to a MoBio PowerSoil DNA Isolation Kit (MoBio, Inc., Carlsbad, CA, USA). DNA was extracted according to the manufacturer's instructions. PCR amplification and sequencing were performed at the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory, in accordance with previously described protocols<sup>52–54</sup>. Briefly, 515F and barcoded 806R primers with Illumina flowcell adapter sequences were used to amplify the V4 region of bacterial and archaeal 16S rRNA genes<sup>55</sup>. Each 25 µl PCR reaction contained 12 µl of PCR water (MoBio, Inc., Carlsbad, CA, USA), 10 µl of 1 × 5 PRIME Hot Master Mix (5 PRIME Inc., Bethesda, MD, USA), 1 µl each of F and R primers (5 µM concentration, 200 pM final), and 1 µl of template DNA. PCR cycling conditions were as follows: 94 °C for 3 min, 35 cycles of [94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s], 72 °C for 10 min. A PicoGreen assay (Life Technologies, Grand Island, NY, USA) was used to measure amplicon concentrations. Equimolar concentrations for each barcoded sample were combined and then cleaned with the UltraClean PCR Clean-Up Kit (MoBio Inc., Carlsbad, CA, USA) and then quantified using the Qubit (Invitrogen, Carlsbad, CA, USA). The pool was then diluted to 2 nM, denatured, and then diluted to a final concentration of 4 pM with a 10% PhiX spike for sequencing on the Illumina MiSeq platform.

**Quantitative PCR (qPCR).** A quantitative polymerase chain reaction (qPCR) was performed to measure microbial abundances in units of 16S rRNA gene copies per g wet sediment<sup>54,56</sup>. Each reaction used 5 µl of 2× SYBR Green PCR Master Mix (Applied Biosystems, Carlsbad, CA, USA), 4 µl of template DNA, and 1 µl of primer mix. The 16S rRNA gene 1406F/1525R primer set (0.4 µM, F-GYACWCACCGCCCGT and R-AAGGAGGTGWTCCARCC) was designed to amplify bacterial and archaeal 16S rRNA genes. The rpsL primer pair (0.2 µM, F-GTAAAGTATGCCGTGTTCTGT and R-AGCCTGCTTACGGTCTTTA) was used for inhibition control samples to amplify *Escherichia coli* DH10B only. Three dilutions (1/100, 1/500, and 1/1000), as well as an inhibition control (1/100 dilution of *E. coli* DH10B genomic DNA spiked into a 1/100 dilution of the sample), were run in triplicate for each sample and standard. For the standards, *E. coli* DH10B genomic DNA dilutions of 10<sup>–2</sup>, 10<sup>–3</sup>, 10<sup>–4</sup>, 10<sup>–5</sup>, and 10<sup>–6</sup> of the 20 ng/µl stock solution were used. The qPCRs were run on the ViiA7 Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA), with cycling conditions as follows: 10 min at 95 °C, 40 cycles of [15 s at 95 °C, then 20 s at 55 °C, then 30 s at 72 °C]. A melt curve was produced by running a cycle of 2 min at 95 °C and a final cycle of 15 s at 60 °C. The cycle threshold (Ct) values were recorded and analyzed using ViiA7 v1.2 software, and 16S rRNA gene copy numbers were calculated for each sample, accounting for the genome size (4,686,137 bp) and 16S rRNA gene copy number (7) of the standard.

**16S rRNA gene sequence processing and OTU table generation.** Sequences were processed as previously described<sup>54</sup>. Briefly, after demultiplexing by sample, each pair of forward and reverse 16S rRNA gene reads was merged. Sequences were then quality-filtered, and singletons were removed with QIIME<sup>57</sup> and UPARSE<sup>58</sup>. Dereplicated sequences were then clustered at 97% nucleotide identity using UCLUST v7<sup>59</sup> to generate a database containing one sequence for each OTU. Sequencing reads from the full dataset were then clustered to the database to generate an OTU table. Each OTU was assigned taxonomy via the Ribosomal Database Project taxonomic classifier<sup>60</sup>, and all OTUs assigned as mitochondria or chloroplasts were removed. The resulting OTU table was rarefied to 3000 16S rRNA gene sequences per sample. Following this OTU table curation, 36 samples across 21 core-depth combinations were retained, of which 30 were replicates (i.e., 15 pairs). For each pair of replicates, each OTU count was averaged (for 14 of 15 pairs, replicates were indistinguishable, Supplementary Fig. 8), and the averages were used for all downstream analyses. For the six samples without successful replicates, OTU counts from a single sample were used.

**Metagenomic sequencing and bioinformatics.** Based on preliminary 16S rRNA gene amplicon sequencing data from eight samples (IHM4, IHM36, IHE4, IHE28, MHM4, MHM34, MHE4, and MHE16), three samples with the most distinct microbial communities (IHM4, IHE28, and MHE16) were selected for metagenomic sequencing to maximize recovery of diverse microbial populations. DNA (from the same extractions described above for 16S rRNA gene sequencing) was sent to the Australian Centre for Ecogenomics for metagenomic library construction and sequencing on the Illumina NextSeq platform, as previously described<sup>36,37</sup>. Metagenomic assembly, genome binning to recover microbial MAGs, and annotation (to predict gene functions and reconstruct metabolic pathways) were performed as previously described<sup>61</sup>. Briefly, each metagenome was separately assembled using the CLC de novo assembler v4.4.1 (CLCBio, Denmark), reads were mapped to contigs using BWA v0.7.12-r1039<sup>62</sup>, and the mean coverage of contigs was obtained using the 'coverage' command of CheckM v1.0.6<sup>63</sup>. Genomes were binned using MetaBAT v0.26.3<sup>64</sup> with all five preset parameters (very sensitive, sensitive, specific, very specific, super specific), and genome completeness and contamination were estimated using CheckM<sup>63</sup>. To investigate predicted metabolic functions of interest in the metagenomic data, metagenomic reads with sequence similarity to genes diagnostic of specific metabolic functions (e.g., methane monooxygenase, *pmoA*, and methyl-coenzyme M reductase, *mcrA*, indicative of aerobic methane oxidation and methanogenesis, respectively) were identified using hidden Markov models via GraftM v1.0<sup>65</sup>.

**Incubations for CH<sub>4</sub> production rates.** Anaerobic incubations of lake sediment samples were performed to assess rates of production of CH<sub>4</sub>. Four replicate sediment samples (4 ml) from three depths in 2012 (0–5, 10, 20 cm) were collected in the field and immediately sealed in a 120 ml serum vial. The headspace was immediately flushed for 5 minutes with Ultra High Purity (UHP) N<sub>2</sub> in the field (replacing the headspace 20 times) to establish an anaerobic headspace. The vials were stored in coolers, taken to the research station, flushed again with N<sub>2</sub> before incubations began in the laboratory, and then stored during the experiment as follows: 2 vials were incubated at 5 °C and 2 vials were held at room temperature (22 °C) for each depth. Consistent with previous incubations from lake sediments<sup>18</sup>, 5 ml of headspace was sampled daily for 5 days and analyzed on a flame ionization GC to determine CH<sub>4</sub> fluxes. Flux was calculated as a change in headspace concentration over time, then normalized by sediment mass after incubations, when vials were dried and weighed to determine sediment dry weight. We also report data from incubations in 2013 that were treated the same way with samples collected at depths consistent with changes in core sediment transitions: Inre Harsrsjön edge: 2.5, 27.5, 47.5 cm; Inre Harsrsjön middle: 4.5, 35, 60 cm; Mellersta Harsrsjön edge: 7.5, 22.5, 37.5 cm; and Mellersta Harsrsjön middle: 2.5, 27.5, 47.5 cm.

**Calculations of depth-resolved fugitive CH<sub>4</sub>.** Depth-resolved fugitive CH<sub>4</sub> (total CH<sub>4</sub> released from the sediments, including ebullitive and diffusive CH<sub>4</sub>, though ebullition is by far the dominant CH<sub>4</sub> production pathway in these sediments, accounting for 80–88% of total emissions<sup>12</sup>) was calculated from the concentration and stable carbon isotopic composition of CH<sub>4</sub> and DIC in sediment porewater<sup>66</sup>. This approach leverages the fact that (1) microbial fermentation and respiration, which generate CO<sub>2</sub>, do not fractionate carbon, while methanogenesis, which generates CH<sub>4</sub> and CO<sub>2</sub> (1:1), does fractionate carbon, and (2) DIC largely remains dissolved in water, while dissolved CH<sub>4</sub> escapes porewater by ebullition. In this framework, the measured isotopic composition of CH<sub>4</sub> in porewater was used to calculate the fractionation factor associated with methanogenesis, assuming the starting isotopic composition of the substrate matched that measured for organic carbon in the sediment. The model assumes that the products of fermentation, including acetate and/or CO<sub>2</sub>, feed directly into methanogenesis as substrates. While evidence for microbial fractionation of the bulk acetate pool has been demonstrated<sup>67</sup>, we do not have knowledge of the isotopic composition of acetate in our system, and thus we could not use this information to constrain our model. The calculated fractionation factor associated with methanogenesis, along with the measured isotopic composition of DIC in porewater, was used to determine the relative amount of DIC that came from methanogenesis vs. non-fractionating pathways (e.g., fermentation). Because any CO<sub>2</sub> produced was assumed to stay dissolved in porewater, the relative amount of DIC generated from methanogenesis could be multiplied by the measured concentration of DIC to determine the concentration of CO<sub>2</sub> and CH<sub>4</sub> generated through methanogenesis. This generated CH<sub>4</sub> concentration was larger than the actual measured concentration of CH<sub>4</sub> in porewater, and the difference between the two was assigned as 'fugitive' CH<sub>4</sub>. Calculations assumed that the system was at a steady state.

**Statistics and reproducibility.** Homogeneity of regression among groups in Arrhenius plots (Fig. 1 and Supplementary Fig. 2) was tested using pairwise full factorial analyses with JMP statistical software (SAS Institute Inc., Cary, NC). The significance level of the group interaction of lake zone by temperature interval was 0.05 (Supplementary Table 3).

Unless otherwise indicated, other statistical analyses were performed using PRIMER v7<sup>68,69</sup>. The rarefied OTU table was square-root transformed, and Bray-Curtis similarity matrices were generated for sample comparisons and used to make a Principal Coordinates Analysis (PCoA) plot. We used permutational

ANOVA (PERMANOVA) to test for significant differences in microbial community composition between categorical groups of samples (e.g., between the two lakes and between the edges and middles of the lakes), and we used Mantel tests with Spearman's rank correlations to compare microbial community composition (Bray–Curtis similarity matrices) to continuous variables (Euclidean distance matrices), including sediment depth and biogeochemical data. ANOVA and linear regression analyses (Supplementary Tables 9 and 11) were performed with StatPlus v6.1.7.0.

We performed PLSR in the R programming language via the package PLS (function `PLSR`)<sup>70–72</sup> to predict measured sediment CH<sub>4</sub> concentrations from biotic and abiotic variables, similar to our previously described PLSR analyses<sup>36</sup>. Briefly, PLSR models a causal relationship between explanatory variable(s) (in this case, abundances of abiotic measurements and/or microorganisms) and the response variable being predicted (here, measured sediment CH<sub>4</sub> concentrations). Abiotic variables included all depth-resolved abiotic measurements that were not directly related to CH<sub>4</sub>, as such measurements could be confounding variables in our analysis. The included abiotic variables were: depth, TOC, δ<sup>13</sup>C<sub>TOC</sub>, DIC, S, and TOC:TS. The PLSR analysis yielded Pearson's product–moment correlations between measured environmental and/or geochemical variables, the abundances of microbial lineages, and the abundances of specific microbial populations. This allowed for a quantification of the added value of microbial abundances in predicting sediment CH<sub>4</sub> concentrations, relative to predictions from abiotic factors alone. Variance in projection (VIP) scores for each explanatory variable indicate the extent to which that variable was predictive of the response variable (i.e., sediment CH<sub>4</sub> concentrations), with VIP scores ≥ 1 considered to be highly significant<sup>73</sup>.

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

### Data availability

Sequencing data are available at NCBI under BioProject PRJNA667178 and also here: [https://isogenie-db.asc.ohio-state.edu/datasources#lake\\_data](https://isogenie-db.asc.ohio-state.edu/datasources#lake_data) (from the IsoGenie link, note that the two folders with MAGs are based on initial taxonomy; some MAGs subsequently determined to be archaea are in the bacteria folder and vice versa). NCBI accession numbers are as follows: raw 16S rRNA gene amplicon sequences SRX10114484–SRX10114504, raw metagenomic sequences SRX10063754–SRX10063756, and MAGs JAFNEO000000000–JAFNIC000000000. Other raw data and relevant processed data are provided in supplementary tables and/or associated with prior publications, as cited in the paper. Data underlying Figs. 1–4 can be found as follows: Fig. 1A–C (Supplementary Table 2), Fig. 2C, D (Supplementary Table 5), Fig. 3 (Supplementary Table 10), and Fig. 4A–C (raw data in Supplementary Tables 4–5, relevant processed data in Supplementary Tables 13 and 16).

Received: 3 March 2020; Accepted: 7 July 2021;

Published online: 05 October 2021

### References

- Wik, M., Varner, R. K., Anthony, K. W., MacIntyre, S. & Bastviken, D. Climate-sensitive northern lakes and ponds are critical components of methane release. *Nat. Geosci.* **9**, 99–105 (2016).
- Wik, M. et al. Energy input is primary controller of methane bubbling in subarctic lakes. *Geophys. Res. Lett.* **41**, 555–560 (2014).
- Holgerson, M. A. & Raymond, P. A. Large contribution to inland water CO<sub>2</sub> and CH<sub>4</sub> emissions from very small ponds. *Nat. Geosci.* **9**, 222–226 (2016).
- Tan, Z. & Zhuang, Q. Arctic lakes are continuous methane sources to the atmosphere under warming conditions. *Environ. Res. Lett.* **10**, 054016 (2015).
- Verpoorter, C., Kutser, T., Seekell, D. A. & Tranvik, L. J. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* **41**, 6396–6402 (2014).
- Walter, K. M., Smith, L. C. & Chapin, F. S. Methane bubbling from northern lakes: present and future contributions to the global methane budget. *Philos. Trans. R. Soc. A* **365**, 1657–1676 (2007).
- van Huissteden, J. et al. Methane emissions from permafrost thaw lakes limited by lake drainage. *Nat. Clim. Change* **1**, 119–123 (2011).
- Schuur, E. A. G. et al. Vulnerability of permafrost carbon to climate change: implications for the global carbon cycle. *BioScience* **58**, 701–714 (2008).
- Wik, M., Crill, P. M., Varner, R. K. & Bastviken, D. Multiyear measurements of ebullitive methane flux from three subarctic lakes. *Biogeochemistry* **118**, 1307–1321 (2013).
- Bastviken, D., Tranvik, L. J., Downing, J. A., Crill, P. M. & Enrich-Prast, A. Freshwater methane emissions offset the continental carbon sink. *Science* **331**, 50 (2011).
- Chanton, J. P. The effect of gas transport on the isotope signature of methane in wetlands. *Org. Geochem.* **36**, 753–768 (2005).
- Jansen, J. et al. Climate-sensitive controls on large spring emissions of CH<sub>4</sub> and CO<sub>2</sub> from northern lakes. *Biogeochemistry* **124**, 2379–2399 (2019).
- Lundin, E. J. et al. Is the subarctic landscape still a carbon sink? Evidence from a detailed catchment balance. *Geophys. Res. Lett.* **43**, 1988–1995 (2016).
- Thornton, B. F., Wik, M. & Crill, P. M. Climate-forced changes in available energy and methane bubbling from subarctic lakes. *Geophys. Res. Lett.* **42**, 1936–1942 (2015).
- Wik, M. et al. Sediment characteristics and methane ebullition in three subarctic lakes. *Biogeochemistry* **123**, 2399–2411 (2018).
- Tranvik, L. J. et al. Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **54**, 2298–2314 (2009).
- Alperin, M. J., Albert, D. B. & Martens, C. S. Seasonal variations in production and consumption rates of dissolved organic carbon in an organic-rich coastal sediment. *Geochim. Cosmochim. Acta* **58**, 4909–4930 (1994).
- Duc, N. T., Crill, P. & Bastviken, D. Implications of temperature and sediment characteristics on methane formation and oxidation in lake sediments. *Biogeochemistry* **100**, 185–196 (2010).
- Winkel, M. et al. First evidence for cold-adapted anaerobic oxidation of methane in deep sediments of thermokarst lakes. *Environ. Res. Commun.* **1**, 021002 (2019).
- Ruuskanen, M. O., St. Pierre, K. A., St. Louis, V. L., Aris-Brosou, S. & Poulain, A. J. Physicochemical drivers of microbial community structure in sediments of Lake Hazen, Nunavut, Canada. *Front. Microbiol.* **9**, 1138 (2018).
- Kokfelt, U. et al. Wetland development, permafrost history and nutrient cycling inferred from late Holocene peat and lake sediment records in subarctic Sweden. *J. Paleolimnol.* **44**, 327–342 (2010).
- Rissanen, A. J. et al. Vertical stratification of bacteria and archaea in sediments of a small boreal humic lake. *FEMS Microbiol. Lett.* **366**, fnz044 (2019).
- Evans, P. N. et al. An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* **17**, 219–232 (2019).
- Parke, R. J. et al. Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**, 390–394 (2005).
- Inagaki, F. et al. Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc. Natl Acad. Sci. USA* **103**, 2815–2820 (2006).
- Colwell, F. S. et al. Estimates of biogenic methane production rates in deep marine sediments at Hydrate Ridge, Cascadia Margin. *Appl. Environ. Microbiol.* **74**, 3444 (2008).
- Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. & House, C. H. Metagenomic signatures of the Peru Margin seafloor biosphere show a genetically distinct environment. *Proc. Natl Acad. Sci. USA* **105**, 10583–10588 (2008).
- Morris, B. E. L., Henneberger, R., Huber, H. & Moissl-Eichinger, C. Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* **37**, 384–406 (2013).
- Kotsyurbenko, O. R. Trophic interactions in the methanogenic microbial community of low-temperature terrestrial ecosystems. *FEMS Microbiol. Ecol.* **53**, 3–13 (2005).
- Vavilin, V., Rytov, S. & Conrad, R. Modelling methane formation in sediments of tropical lakes focusing on syntrophic acetate oxidation: dynamic and static carbon isotope equations. *Ecol. Model.* **363**, 81–95 (2017).
- Liu, P. & Conrad, R. Syntrophobacteraceae-affiliated species are major propionate-degrading sulfate reducers in paddy soil. *Environ. Microbiol.* **19**, 1669–1686 (2017).
- Sieber, J. R., McInerney, M. J. & Gunsalus, R. P. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Ann. Rev. Microbiol.* **66**, 429–452 (2012).
- Bhattarai, S., Zhang, Y. & Lens, P. N. L. Effect of pressure and temperature on anaerobic methanotrophic activities of a highly enriched ANME-2a community. *Environ. Sci. Pollut. Res.* **25**, 30031–30043 (2018).
- Ruff, S. E. et al. Global dispersion and local diversification of the methane seep microbiome. *Proc. Natl Acad. Sci. USA* **112**, 4015–4020 (2015).
- McCalley, C. K. et al. Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478–481 (2014).
- Emerson, J. B. et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).
- Woodcroft, B. J. et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
- Johnston, E. R. et al. Responses of tundra soil microbial communities to half a decade of experimental warming at two critical depths. *Proc. Natl Acad. Sci. USA* **116**, 15096–15105 (2019).
- Zinke, L. A. et al. Microbial organic matter degradation potential in Baltic Sea sediments is influenced by depositional conditions and in situ geochemistry. *Appl. Environ. Microbiol.* **85**, e02164–02118 (2019).
- Sharon, I. et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res.* **25**, 534–543 (2015).
- Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Karnachuk, O. V. & Ravin, N. V. Genome of the candidate phylum Aminicenantes bacterium from a deep

- subsurface thermal aquifer revealed its fermentative saccharolytic lifestyle. *Extremophiles* **23**, 189–200 (2019).
42. Robbins, S. J., Evans, P. N., Parks, D. H., Golding, S. D. & Tyson, G. W. Genome-centric analysis of microbial populations enriched by hydraulic fracture fluid additives in a coal bed methane production well. *Front. Microbiol.* **7**, 731 (2016).
  43. Gies, E. A., Konwar, K. M., Beatty, J. T. & Hallam, S. J. Illuminating microbial dark matter in meromictic Sakinaw Lake. *Appl. Environ. Microbiol.* **80**, 6807 (2014).
  44. Lloyd, K. G. et al. Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496**, 215–218 (2013).
  45. Spring, S., Bunk, B., Spröer, C., Rohde, M. & Klenk, H.-P. Genome biology of a novel lineage of planctomycetes widespread in anoxic aquatic environments. *Environ. Microbiol.* **20**, 2438–2455 (2018).
  46. Borrel, G. et al. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. *BMC Genomics* **15**, 679 (2014).
  47. Vanwonterghem, I. et al. Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
  48. Zinke, L. A. et al. Evidence for non-methanogenic metabolisms in globally distributed archaeal clades basal to the Methanomassiliicoccales. *Environ. Microbiol.* **23**, 340–357 (2021).
  49. Wik, M., Crill, P. M., Bastviken, D., Danielsson, Å. & Norbäck, E. Bubbles trapped in arctic lake ice: potential implications for methane emissions. *Biogeoscience* **116**, G03044 (2011).
  50. Magen, C. et al. A simple headspace equilibration method for measuring dissolved methane. *Limnol. Oceanogr. Methods* **12**, 637–650 (2014).
  51. Pimmel, A. & Claypool, G. Introduction to shipboard organic chemistry on the JOIDES Resolution. *ODP Tech. Note* **30**, 1–29 (2001).
  52. Bates, S. T. et al. Examining the global distribution of dominant archaeal populations in soil. *ISME J.* **5**, 908–917 (2011).
  53. Lauber, C. L., Ramirez, K. S., Aanderud, Z., Lennon, J. & Fierer, N. Temporal variability in soil microbial communities across land-use types. *ISME J.* **7**, 1641–1650 (2013).
  54. Emerson, J. B. et al. Impacts of flood damage on airborne bacteria and fungi in homes after the 2013 Colorado Front Range Flood. *Environ. Sci. Technol.* **49**, 2675–2684 (2015).
  55. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2010).
  56. Emerson, J. B. et al. High temporal variability in airborne bacterial diversity and abundance inside single-family residences. *Indoor Air* **27**, 576–586 (2017).
  57. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
  58. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
  59. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
  60. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
  61. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
  62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
  63. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
  64. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
  65. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucl. Acid. Res.* **46**, e59–e59 (2018).
  66. Corbett, J. E. et al. Partitioning pathways of CO<sub>2</sub> production in peatlands with stable carbon isotopes. *Biogeochem* **114**, 327–340 (2013).
  67. Heuer, V. B., Pohlman, J. W., Torres, M. E., Elvert, M. & Hinrichs, K.-U. The stable carbon isotope biogeochemistry of acetate and other dissolved carbon species in deep seafloor sediments at the northern Cascadia Margin. *Geochim. Cosmochim. Acta* **73**, 3323–3336 (2009).
  68. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
  69. Clarke, K. R. & Gorley, R. N. *PRIMER version 6: user manual/tutorial* (PRIMER-E, Plymouth, 2006).
  70. Lê Cao, K.-A., Rossouw, D., Robert-Granié, C. & Besse, P. A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* **7**, 35 (2008).
  71. Shen, H. & Huang, J. Z. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* **99**, 1015–1034 (2008).
  72. Guidi, L. et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016).
  73. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemometr. Intell. Lab. Syst.* **78**, 103–112 (2005).

## Acknowledgements

We would like to acknowledge the following funding in support of this project: U.S. Department of Energy grants (DE-SC0010580 and DE-SC0016440, Co-lead PI Rich; DE-SC0010338 and DE-SC0019063, PI Neumann; DE-SC0020163, Co-PI Emerson), U.S. National Science Foundation grants (the Northern Ecosystems Research for Undergraduates program (NERU) EAR-1063037, PI Varner, MacroSystems Biology grant EF #1241037, PI Varner, and the EMERGE Biology Integration Institute grant #2022070, PI Rich), the Swedish Research Council grants to P. Crill (2007-4547 and 2013-5562), and the UC Davis College of Agricultural and Environmental Sciences and Department of Plant Pathology (laboratory start-up funds to JBE). Thanks to staff at the Polar Research Secretariat's Abisko Research Station (ANS). Thanks to Kaitlyn Steele, Florencia Fahnestock, Kiley Remiszewski, Carmody McCalley, and NERU participants Sophia Burke, Joel DeStasio, Lance Erickson, and Madison Halloran for assistance in sample collection and analysis, Jacob Setera and Steve Phillips (UNH) for assistance with the CHNS elemental analysis, Peter Sternes for assistance with data upload to NCBI, Joachim Jansen for helping us to think critically about the ebullition results, and the IsoGenie Project Team for discussion of results.

## Author contributions

J.B.E., R.K.V., and V.I.R. designed the study, J.B.E. wrote and R.K.V. and V.I.R. substantially revised the paper. R.K.V., M.W., and N.L.F. collected the samples and J.B.E., R.K.V., M.W., J.E.J., A.O.D., M.B., and N.L.F. performed laboratory sample processing. M.W., R.K.V., and P.M.C. generated and analyzed ebullitive CH<sub>4</sub> flux data, R.K.V. and J.E.J. generated and analyzed geochemical data, R.K.V. and N.L.F. performed and analyzed ex situ incubations, R.B.N. performed the isotopic and mass balance calculations, J.B.E., D.H.P., C.M.S., and B.J.W. processed and J.B.E. and R.T. II analyzed microbial community sequencing data, J.B.E. reconstructed microbial metabolic pathways with guidance from D.H.P., C.M.S., B.J.W., and G.W.T., and J.B.E. performed the statistical modeling with guidance from S.R.S. J.B.E., R.K.V., M.W., D.H.P., R.B.N., J.E.J., C.M.S., B.J.W., R.T. II, N.L.F., A.O.D., M.B., and V.I.R. performed the data analysis. All authors contributed to project discussions, edited the paper, and approved the final version of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information


**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25983-9>.

**Correspondence** and requests for materials should be addressed to Joanne B. Emerson, Ruth K. Varner or Virginia I. Rich.

**Peer review information** *Nature Communications* thanks Stephanie Carr, Florence Schubotz, and Cornelia Welte for their contributions to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021