# Genome-centric view of carbon processing in thawing permafrost

Ben J. Woodcroft[1*], Caitlin M. Singleton[1*], Joel A. Boyd[1], Paul N. Evans[1], Joanne B. Emerson[2†], Ahmed AF. Zayed[2], Robert D. Hoelzle[1], Timothy O. Lamberton[1], Carmody K. McCalley[3], Suzanne B. Hodgkins[4], Rachel M. Wilson[4], Samuel O. Purvine[5], Carrie D. Nicora[5], Changsheng Li[6], Steve Frolking[6], Jeff P. Chanton[4], Patrick M. Crill[7], Scott R. Saleska[8], Virginia I. Rich[2], Gene W. Tyson[1#]

[1] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, Australia

[2] Department of Microbiology, The Ohio State University, Columbus, OH 43210, United States

[3] Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY 14623, United States

[4] Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32306, United States

[5] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, United States

[6] Earth Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire, United States

[7] Department of Geological Sciences and Bolin Centre for Climate Research, Stockholm University, 10691 Stockholm, Sweden

[8] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85716, United States

[*] These authors contributed equally to this work

[†] Current address: Department of Plant Pathology, University of California, Davis, CA 95616, United States

[#] Corresponding author: Prof. Gene W. Tyson. Mailing address: Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD 4072, Australia. Phone: +617 3365 3829 Fax: +617 336 54511 Email: g.tyson@uq.edu.au

**As global temperatures rise, large amounts of carbon sequestered in permafrost are becoming available for microbial degradation. Accurate prediction of carbon gas emissions produced from thawing permafrost is limited by our understanding of these microbial communities. Here, metagenomic sequencing of 214 samples from a permafrost thaw gradient enabled the recovery of 1,529 metagenome-assembled genomes, including many from phyla with poor genomic representation. These genomes were shown to reflect the diversity of this complex ecosystem with genus-level representatives for >60% of the community. Meta-omic analysis revealed key populations involved in organic matter degradation, including bacteria encoding a novel fungal pathway for xylose degradation. Microbial and geochemical data highlighted lineages correlated with the production of greenhouse gases and indicated novel syntrophic relationships. Our findings link changing biogeochemistry to specific microbial lineages involved in carbon processing, providing key information for predicting the impact of climate change on permafrost systems.**

Climate change-induced permafrost thaw is predicted to make up to 174 Pg of near surface carbon (<3 m) available for microbial degradation by 2100[1]. Predicting the magnitude of carbon loss as carbon dioxide ($CO_2$) or methane ($CH_4$) is hampered by our limited knowledge of microbial organic matter (OM)

metabolism in these environments. Genome-centric metagenomic analysis of microbial communities provides the necessary information to examine how specific lineages transform OM during permafrost thaw. However, these methods are challenged by the inherent complexity and spatial heterogeneity of near-surface soil communities that support diverse functional processes[2-4]. Prior metagenomic studies in permafrost-associated soils from Alaskan tundra[5,6] and a mineral soil permafrost[7] recovered a small number of metagenome-assembled genomes (14-33 MAGs), which represent only a fraction of the species present in these systems. However, the ability to recover MAGs from complex microbial communities is continually improving in parallel with advances in sequencing technology and bioinformatic techniques[8,9].

**Recovery and distribution of MAGs**

The discontinuous permafrost at Stordalen Mire in northern Sweden is a model Arctic peatland ecosystem for studying thaw progression[10]. To gain an understanding of the microbial communities and their associated carbon metabolism, soil samples were collected from the active layer (seasonally thawed) of three sites across a thaw gradient: an intact palsa (thawed to ~30 cm), a partially-thawed bog (~60 cm), and a fully-thawed fen. While bogs and fens exist in diverse landscapes, thaw-associated shifts in hydrology cause them to be a common feature of thawing northern peatland permafrost systems (see **Methods**). Triplicate soil cores and biogeochemical measurements were taken at each site from four active layer depths (near surface, mid, deep and extra-deep; 1-51 cm; **Fig. 1** & **Supplementary Data File 1**) over three growing seasons. In total, 1.7 Tbp of metagenomic sequence data was generated from 214 samples (2-165 Gbp per sample), with supporting metatranscriptomic and metaproteomic data from a subset of these samples (**Supplementary Data File 1, 2**). Metagenome assembly and differential coverage binning yielded 1,529 medium to high-quality MAGs (>70% complete and <10% contaminated) from a diverse range of bacterial (1,434 genomes) and archaeal phyla (95 genomes; **Extended Data Fig. 1a, Supplementary Data File 3**). The Stordalen Mire MAGs expand the number of genomes recovered from permafrost-associated soils by two orders of magnitude.

To resolve the taxonomic distribution of the Stordalen Mire MAGs, phylogenetic trees were inferred from concatenated sets of single-copy marker genes (120 bacterial or 122 archaeal genes). The recovered MAGs spanned 30 phyla, including Bacteria belonging to Actinobacteria (385 genomes), Acidobacteria (364), Proteobacteria (205), Chloroflexi (66) and Archaea from the Euryarchaeota (85). The Stordalen genomes substantially expand representation of several common soil dwelling lineages (**Extended Data Fig. 1**), such as the ubiquitous Acidobacteria, where genomic representation was increased three fold. MAGs were also recovered from many poorly characterized phyla, including 47 genomes from the

bacterial candidate phylum Dormibacteraeota (AD3), 53 from Eremiobacteraeota (WPS-2), six from FCPU426 and eight archaea from the Bathyarchaeota (**Extended Data Fig. 1, Supplementary Data File 10, 11**). The Stordalen genomes broadly represent the major groups present in the system (**Extended Data Fig. 2a & b**) as well as many lineages previously detected in other permafrost-associated environments[6,11]. Based on the diversity of ribosomal protein sequences detected in the metagenomes, we conservatively estimate that there are >24,000 strains inhabiting Stordalen Mire (**Supplementary Note 1**). The Stordalen MAGs represent ~60% of microorganisms in the mire at the genus level (**Supplementary Note 2**), making this the most comprehensive recovery of genomes from a complex, natural soil environment to date.

Stordalen genomes were explicitly linked to the changing habitats by mapping their abundances across the thaw gradient (**Fig. 1, Supplementary Data File 4**). Communities shifted substantially between sites (**Extended Data Fig. 2c**), with MAGs belonging to the Acidobacteria, Actinobacteria, Eremiobacteraeota, Alpha- and Gammaproteobacteria predominant in the palsa and bog (5 - 41% of the community), while Deltaproteobacteria, Bacteroidetes, Chloroflexi and Ignavibacteriae were almost exclusively observed in the fen (8 - 14%; **Extended Data Fig. 2e & f**). The extra-deep bog samples had the lowest diversity (Shannon index 3.74±0.24; **Fig. 1**), potentially due to the ombrotrophic and anaerobic conditions in this environment, whereas the shallow samples from the minerotrophic fen were the most diverse (Shannon index 4.55±0.05; **Fig. 1**). The fen also had 2.6× more microbial cells per gram of soil relative to the palsa and bog (**Extended Data Fig. 2d**). In the bog, decreasing oxygen availability with depth likely drives changes in community structure, with Euryarchaeota and Dormibacteraeota increasing in relative abundance with depth, and Eremiobacteraeota decreasing. In the fen, the Planctomycetes, Omnitrophica and Spirochaetes increased in abundance with depth (**Extended Data Fig. 2e & f**). Consistent with the heterogeneity of soil environments[12,13], individual MAGs were typically only found at high abundance (>1%) in a limited number of samples (<4). However, a small number of MAGs belonging to the Acidobacteria, Actinobacteria, Proteobacteria and Euryarchaeota were ubiquitous at specific depths within the palsa, bog or fen (**Extended Data Fig. 3a & b**; **Extended Data Table 1**). For several genera, closely related MAGs were abundant at different depths (**Extended Data Fig. 3c**), reflecting fine-scale adaptation to distinct niches in the soil column.

**Polysaccharide degradation**

Metabolic reconstruction of the MAGs, combined with 24 metatranscriptome and 16 metaproteome datasets (**Extended Data Table 2**), allowed examination of the key microorganisms, pathways and interactions responsible for OM degradation and the production of greenhouse gases at Stordalen Mire

(**Fig. 2; Supplementary Data File 1**). The first stage in degradation involves the breakdown of high molecular weight plant-derived polysaccharides, primarily cellulose and hemicellulose, which make up a large proportion of peatland carbon[14]. The ability to degrade these polysaccharides was a dominant feature of the Stordalen MAGs across all three thaw environments (**Supplementary Data File 5**; **Fig. 2** 'MAG abundances' & 'distribution boxplots'), with many encoding cellulases and xylanases (39% and 37%, respectively, average 3.8 and 2.6 copies per genome). This is consistent with gene-centric metagenomic studies of Arctic fens and tundras[15,16], however, the genome-centric approach used here links these metabolic functions to specific populations.

Cellulase- and xylanase-encoding microorganisms, primarily belonging to the Acidobacteria (**Fig. 2** 'MAG abundances'), were most abundant in the palsa surface (68% and 59% of the recovered community, respectively), and decreased with depth (**Fig. 2 '**distribution boxplots'). The surface bog had the lowest percentage of microorganisms encoding these genes (34% and 24%, respectively), likely due to breakdown inhibition through the production of acids by *Sphagnum* moss[17]. The high relative abundance of cellulase and xylanase-encoding Acidobacteria (61% and 75% of acidobacterial genomes, respectively), strongly suggests that they are the primary degraders of large polysaccharides in the palsa and bog (**Figs. 1 & 2**). Metatranscriptomic data confirmed these genomic inferences, with Acidobacteria producing the majority of cellulase and xylanase transcripts at these sites (**Extended Data Fig. 4**). Metaproteomic analysis revealed protein expression for 45 cellulases and 27 xylanases primarily belonging to Acidobacteria in the bog (**Supplementary Data File 2**). A wider range of microorganisms are responsible for this functionality in the fen, including members of the Proteobacteria, Ignavibacteriae, Bacteroidetes, Verrucomicrobia, Chloroflexi, and Actinobacteria. However, the metatranscriptomic data indicates the Proteobacteria, Ignavibacteriae and Bacteroidetes have the highest expression of these genes in the fen, although only a limited number of proteins were detected. Notably most putative cellulose hydrolyzers also encode a xylanase (59% of genomes), with the exception of actinobacterial hydrolyzers, which typically only encode cellulases (87%). However, unlike findings in other Arctic systems[18], both metatranscriptomic and metaproteomic data show that the actinobacterial cellulases are not highly expressed, indicating that these microorganisms play a minor role in polysaccharide degradation at Stordalen Mire. The high abundance of hydrolyzers in the palsa suggests that the microbial community contributes to physical compaction through decomposition of surface organic matter, as evidenced by increases in bulk density with depth (**Supplementary Note 3**), and that this contribution likely augments thawing of underlying permafrost as the primary driver of subsidence.

Breakdown of polysaccharides into simple sugars is the primary source of energy and carbon for the microbial community[16]. β-glucosidases for disaccharide degradation were encoded by the majority of

microorganisms in all sites (75%, 84% and 66% of palsa, bog and fen communities), with transcript expression primarily by Acidobacteria in the palsa and bog, and Bacteroidetes in the fen (**Extended Data Fig. 4**). The metaproteomes supported the high expression of β-glucosidase proteins by Acidobacteria in the bog (198 out of a total of 216 detected proteins). Degradation pathways for the monosaccharides glucose, galactose and xylose were also prevalent in the MAGs (**Supplementary Note 4**, **Extended Data Fig. 5**). Of the 237 microorganisms potentially capable of xylan degradation, 108 appear to be involved in xylose degradation using the canonical bacterial isomerase pathway[19] (**Extended Data Fig. 6a)**. These genomes were common in the surface palsa (40% of the microbial community), deep bog (49%) and throughout the fen (51%), similar to the distribution of microorganisms capable of degrading the precursor xylan. Transcription of this pathway was highest in the fen by Bacteroidetes and Ignavibacteriae (**Fig. 2** 'pathway expression'; **Extended Data Fig. 6**). Members of the Acidobacteria, Actinobacteria and Verrucomicrobia showed highest transcription of this pathway in the bog, while expression was limited to Actinobacteria in the palsa (**Extended Data Fig. 6c-h**). Intriguingly, 50 actinobacterial MAGs encoded genes necessary for xylose degradation, despite 44 being unable to degrade the precursor xylan, indicating that they are reliant on the activity of xylan hydrolyzers.

Only a small fraction of the dominant acidobacterial xylan hydrolyzers encode the necessary genes for the canonical isomerase pathway for xylose degradation (30 of 111 genomes). Few acidobacterial genomes (23) encoded the alternative xylonate dehydratase pathway, indicating they may degrade xylose through a membrane-bound glucose dehydrogenase as previously observed only in *Gluconobacter oxydans*[20] (**Supplementary Note 5**). A closer inspection of acidobacterial xylan hydrolyzing MAGs revealed an oxidoreductase pathway for the conversion of xylose into xylulose only previously identified in fungi (37 genomes; **Supplementary Note 5**)[21-23]. MAGs belonging to the Actinobacteria and Chloroflexi also encoded this pathway, together comprising 13% of the community across the thaw gradient. Acidobacterial and actinobacterial genes for the oxidoreductase pathway were expressed in metatranscriptomes from across the mire, and were more highly expressed than the canonical isomerase pathway in both the palsa and bog (**Fig. 2 'pathway expression', Extended Data Fig. 6c-h**). Nine MAGs expressed proteins for this pathway, primarily in the bog, confirming that this novel pathway is in use and likely accounts for a substantial fraction of xylose degradation at the mire (**Supplementary Data File 2**). The detection and expression of several distinct pathways for xylose degradation, often occurring in the same genome (**Extended Data Fig. 6b**), reveals that one or multiple pathways may be active under specific environmental conditions (**Supplementary Note 5**).

**Fermentation**

In the anaerobic layers of the peat column, where inorganic terminal electron acceptors (TEAs) are rare[24,25], fermentation and acetogenesis are essential pathways for the further degradation of monosaccharides, and supply the substrates for methanogenesis. Fermentation produces low molecular weight alcohols and organic acids such as ethanol, propionate, acetate and lactate, as well as hydrogen and $CO_2$[24,26]. In the palsa and bog, lactate fermentation is a common metabolism encoded by actinobacterial and acidobacterial MAGs (**Fig. 2** 'MAG abundances'), which are particularly abundant in the bog surface (36% and 16% of the community respectively), but decrease with depth (**Fig. 2** 'distribution boxplots'). Transcript expression of this pathway, while low across all sites**)**, appears to be mostly limited to these lineages in the bog (**Extended Data Fig. 7**). Conversely, populations belonging to Chloroflexi, Ignavibacteriae, Bacteriodetes, and Proteobacteria appear to be the primary lactate metabolisers in the fen (9%, 7%, 5% and 5% respectively), with Ignavibacteriae the most transcriptionally active (**Extended Data Fig. 7**). A small fraction of Stordalen genomes are capable of ethanol and propionate fermentation (**Fig. 2** 'MAG abundances'), and expression of these pathways is low and mainly limited to the palsa and bog (**Extended Data Fig. 7**). The abundance of these microorganisms in the palsa suggests that they are potentially important fermenters during the early stages of thaw (**Fig. 2** 'distribution boxplots'). Acetogens were most abundant in the fen across all depths, which suggests increased acetate production and is consistent with a preference for pH neutral environments[24,25] (**Fig. 2** 'distribution boxplots'). Fen acetogens belong to the Ignavibacteriae, Bacteroidetes and Verrucomicrobia, whereas in the palsa and bog this metabolism was limited to a small number of Acidobacteria, Actinobacteria and Verrucomicrobia (**Fig. 2** 'MAG abundances'). These distributions were also observed in the metatranscriptomes (**Extended Data Fig. 7**), with Bacteroidetes, Ignavibacteriae and Proteobacteria contributing to the slightly higher expression of acetogenesis transcripts in the fen, compared to expression by Acidobacteria and Verrucomicrobia in the bog (**Extended Data Fig. 7**). While acetate can be oxidised using available inorganic TEAs[27] (e.g. sulphate, nitrate), these are at very low concentrations in Stordalen Mire (**Supplementary Data File 6**). The unexpectedly high ratio of $CO_2$ to $CH_4$ produced at the site[28] (16:1 in the bog and 7:1 in the fen; **Extended Data Fig. 8a**) may signal the oxidation of fermentation products including acetate using organic TEAs such as humic substances[27,29,30].

**Methane metabolism**

Methanogenesis is the final step in anaerobic carbon transformation and is of critical concern in thawing permafrost peatland systems where $CH_4$ release is rapidly increasing[31]. Of the 95 archaeal genomes recovered **(Extended Data Fig. 1)**, 76 were identified as hydrogenotrophic methanogens ($H_2$- and $CO_2$-utilising), which alongside high transcript and protein expression **(Extended Data Fig. 8d & e)** suggests this is the dominant form of $CH_4$ generation at the mire. Hydrogenotrophic methanogens increased in

abundance, diversity and activity as thaw progressed from bog to fen (**Fig. 2** 'MAG abundances'), consistent with the increase in $CH_4$ flux[10], with mid and deep fen samples having the highest relative abundance of these methanogens **(Fig. 2** 'distribution boxplots'**)**. Only six low-abundance acetoclastic methanogens were recovered, primarily from the fen where acetogenesis was also prevalent **(Fig. 2** 'MAG abundances'**)**. Additionally, two $H_2$-dependent methylotrophic methanogens from the order *Methanomassiliicoccales* were recovered, but were present at very low abundance (0.1% in the fen) with low transcriptional activity, making it unlikely that they contribute substantially to $CH_4$ production at the mire (**Extended Data Figure 8d**). Methanotrophs from the Alpha- and Gammaproteobacteria were identified across the thaw gradient. High abundances in the bog suggests methanotrophs may oxidise significant proportions of $CH_4$, limiting emissions to the atmosphere (**Supplementary Note 6**).

**Microbial and geochemical interactions**

The activity of methanogens and methanotrophs alters the $^{13}C/^{12}C$ isotopic ratio of $CH_4$ dissolved in the porewater[32]. A previous 16S rRNA gene amplicon survey at Stordalen Mire revealed that the abundance of *Ca.* 'Methanoflorens stordalenmirensis' was the best single predictor of carbon isotopic fractionation during $CH_4$ production at the bog in 2011[33,34]. The recovery of 51 additional genomes here greatly expands the representation of the order *Ca*. 'Methanoflorentales', and revealed the presence of two habitat-specific clades derived from the bog and fen, respectively (80-85% AAI; *Ca.* 'M. stordalenmirensis' from bog and *Candidatus* 'Methanoflorens crillii' from the fen). The 16S rRNA gene-based correlation of *Ca*. 'M. storadalenmirensis' to the isotopic signature of $CH_4$ was confirmed by the metagenomic data for both 2011 and 2012, with the relative abundance of 19 *Ca.* 'M. stordalenmirensis' MAGs in the bog explaining more variation than bulk environmental variables (2011 $R^2$ 0.43, p-value $6\times10^{-4}$; 2012 $R^2$ 0.48, p-value $2\times10^{-4}$; **Extended Data Fig. 8c & e**). Intriguingly, the relative abundance of a previously uncharacterised acidobacterial population, *Candidatus* 'Acidiflorens stordalenmirensis' was significantly correlated with the isotopic composition of $CH_4$ ($R^2$ 0.40, p-value $2\times10^{-6}$), and even more strongly correlated with the relative abundance of *Ca.* 'M. stordalenmirensis' ($R^2$ 0.82, p-value $< 2\times10^{-16}$ in bog sites; **Fig. 3**, **Supplementary Note 7**).

A detailed metabolic analysis of *Ca.* 'A. stordalenmirensis' and the 49 other genomes belonging to the genus *Ca*. 'Acidiflorens' revealed metabolic capabilities that suggest members of this lineage are facultative syntrophs (**Extended Data Fig. 1b**). Members of *Ca*. 'Acidiflorens' contain genes for the fermentation of a wide range of substrates including xylan, fatty acids, oxalate and fructose, and encode numerous hydrogenases indicating the potential for $H_2$ production and consumption (**Fig. 3d**, **Supplementary Note 8**). We hypothesise that the correlation between *Ca.* 'A. stordalenmirensis' and *Ca.*

'M. stordalenmirensis' and the $CH_4$ isotopic composition indicates that these lineages are in a syntrophic relationship based on inter-species hydrogen transfer. Hydrogen consumption by Ca. 'M. stordalenmirensis' likely lowers the hydrogen partial pressure making fermentation more thermodynamically favourable for *Ca.* 'A. stordalenmirensis'[35,36] (**Fig. 3d**). This syntrophy is also observed in the fen sites between closely related populations, as the relative abundances of a *Ca.* 'Methanoflorens' species (*Ca.* 'M. crillii'; see above) and a second *Ca.* 'Acidiflorens' species (*Ca.* 'Acidiflorens' sp. 2), both of which are only detected in the fen, were also correlated ($R^2$ 0.19, p-value $10^{-4}$; **Fig. 3c**). The species-level resolution afforded by the genome-centric metagenomic approach allowed identification of potential interactions between microorganisms and biogeochemistry that would have been missed using traditional gene amplicon surveys.

Another key but poorly constrained biogeochemical parameter in global $CH_4$ models is the percentage of carbon mineralized to $CO_2$ versus $CH_4$[37]. We directly examined the relationship of microbial lineages in the bog with the porewater $CH_4$:$CO_2$ and identified a significant positive correlation with a genus within the candidate phylum Dormibacteraeota[38], named here *Candidatus* 'Changshengia' ($R^2$ 0.19, p-value 0.001; **Extended Data Fig. 9a & c**). *Ca.* 'Changshengia' was found to oxidise glycerol, an important cryoprotectant in this Arctic environment[39,40], and its derivatives glycerol 3-phosphate and dihydroxyacetone. Based on the transcript and protein expression of genes for glycerol oxidation (**Extended Data Fig. 9b, Supplementary Note 9**), it is possible that *Ca*. 'Changshengia' ferments glycerol leading to the production of $H_2$[41], which is transferred to methanogens, increasing the $CH_4$ to $CO_2$ ratio in the porewater.

**Conclusion**

Here, genome-centric metagenomic analysis of a permafrost thaw gradient allowed the recovery of 1,529 MAGs, significantly increasing the number of genomes sequenced from permafrost-associated environments. Analysis of these genomes, their abundances and expression enabled correlations between specific microbial populations and biogeochemistry to be identified, and revealed key populations driving the mineralisation of OM from plant-derived polysaccharides through to simple sugars, and the greenhouse gases $CO_2$ and $CH_4$. Future efforts combining genome-centric meta-omic data in combination with metabolomics and biogeochemical data will further improve our understanding of large-scale complex global processes, and inform Earth-system models for accurate predictions of climate-induced change.

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

**Acknowledgements** This study was funded by the Genomic Science Program of the United States Department of Energy (DOE) Office of Biological and Environmental Research (BER), grants DE-SC0004632, DE-SC0010580 and DE-SC0016440. B.J.W. and P.N.E. are supported by Australian Research Council Discovery Early Career Research Awards #DE160100248 and #DE170100428, respectively. C.M.S. and J.A.B. are supported by the Australian Government Research Training Program (RTP) Scholarship, and G.W.T. is supported by the Australian Research Council Future Fellowship FT170100070. A portion of the research was performed using Environmental Molecular Sciences Laboratory (EMSL), a DOE Office of Science User Facility, and a portion was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative with resources at both the DOE Joint Genome Institute and EMSL. Both facilities are sponsored by the Office of BER and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). We thank the IsoGenie 1 and 2 Project Teams and the 2010-2012 field teams for sample collection, particularly Tyler Logan, as well as the Abisko Scientific Research Station for sampling infrastructure and support. We thank Philip Hugenholtz, Donovan Parks, Steven Robbins, Brian Kemish, Maria Chuvochina, Serene Low and Margaret Butler for helpful discussion and infrastructure support.

**Author Contributions** C.L., S.F., J.P.C., P.M.C., S.R.S., V.I.R. and G.W.T. designed the overall study of microbial and biogeochemical dynamics of permafrost thaw and procured funding. V.I.R. coordinated sampling efforts, and S.B.H., J.P.C. and G.W.T. collected samples. B.J.W., C.M.S., J.A.B., V.I.R. and G.W.T. designed experiments around specific microbial hypotheses. J.B.E., A.A.Z., S.O.P., and C.D.N performed the protein extractions and analyses. C.K.M., S.B.H. and R.M.W. performed geochemical analyses. B.J.W., C.M.S., J.A.B., P.N.E, R.D.H. and T.O.L. carried out microbial experiments and integration of microbial and geochemical data. B.J.W., C.M.S. and G.W.T. wrote the manuscript with contributions from J.A.B., P.N.E. and R.D.H. All authors excepting C.L. edited, reviewed and approved the final manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.W.T (gene.tyson@uq.edu.au).

## REFERENCES

1 Schuur, E. *et al.* Climate change and the permafrost carbon feedback. *Nature* **520**, 171-179 (2015).

2 Roesch, L. F. *et al.* Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal* **1**, 283-290 (2007).

3 Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature communications* **7** (2016).
4 Howe, A. C. *et al.* Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences* **111**, 4904-4909 (2014).
5 Johnston, E. R. *et al.* Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. *Frontiers in microbiology* **7** (2016).
6 Taş, N. *et al.* Landscape topography structures the soil microbiome in arctic polygonal tundra. *Nature communications* **9**, 777, doi:10.1038/s41467-018-03089-z (2018).
7 Hultman, J. *et al.* Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* (2015).
8 Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature biotechnology* **31**, 533-538 (2013).
9 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7** (2016).
10 Johansson, T. *et al.* Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Global Change Biology* **12**, 2352-2369 (2006).
11 Jansson, J. K. & Taş, N. The microbial ecology of permafrost. *Nature reviews Microbiology* **12**, 414-425 (2014).
12 Whalen, S. C. Biogeochemistry of methane exchange between natural wetlands and the atmosphere. *Environmental Engineering Science* **22**, 73-94 (2005).
13 Rosenzweig, M. L. *Species diversity in space and time*. (Cambridge University Press, 1995).
14 Kremer, C., Pettolino, F., Bacic, A. & Drinnan, A. Distribution of cell wall components in Sphagnum hyaline cells and in liverwort and hornwort elaters. *Planta* **219**, 1023-1035 (2004).
15 Tveit, A., Schwacke, R., Svenning, M. M. & Urich, T. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *The ISME journal* **7**, 299-311 (2013).
16 Ivanova, A. A., Wegner, C. E., Kim, Y., Liesack, W. & Dedysh, S. N. Identification of microbial populations driving biopolymer degradation in acidic peatlands by metatranscriptomic analysis. *Molecular Ecology* **25**, 4818-4835 (2016).
17 Pankratov, T. A., Ivanova, A. O., Dedysh, S. N. & Liesack, W. Bacterial populations and environmental factors controlling cellulose degradation in an acidic Sphagnum peat. *Environmental microbiology* **13**, 1800-1814 (2011).
18 Tveit, A., Urich, T. & Svenning, M. M. Metatranscriptomic analysis of Arctic peat soil microbiota. *Applied and environmental microbiology*, AEM. 01030-01014 (2014).
19 Jeffries, T. W. in *Pentoses and Lignin* 1-32 (Springer, 1983).
20 Zhang, M. *et al.* Genetic analysis of D-xylose metabolism pathways in Gluconobacter oxydans 621H. *Journal of Industrial Microbiology & Biotechnology* **40**, 379-388 (2013).
21 Kricka, W., Fitzpatrick, J. & Bond, U. Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective. *Frontiers in microbiology* **5**, 174 (2014).
22 Kuhn, A., van Zyl, C., van Tonder, A. & Prior, B. A. Purification and partial characterization of an aldo-keto reductase from Saccharomyces cerevisiae. *Applied and environmental microbiology* **61**, 1580-1585 (1995).
23 Sarthy, A. V., Schopp, C. & Idler, K. B. Cloning and sequence determination of the gene encoding sorbitol dehydrogenase from Saccharomyces cerevisiae. *Gene* **140**, 121-126 (1994).
24 Ye, R. *et al.* pH controls over anaerobic carbon mineralization, the efficiency of methane production, and methanogenic pathways in peatlands across an ombrotrophic–minerotrophic gradient. *Soil Biology and Biochemistry* **54**, 36-47 (2012).
25 Horn, M. A., Matthies, C., Küsel, K., Schramm, A. & Drake, H. L. Hydrogenotrophic methanogenesis by moderately acid-tolerant methanogens of a methane-emitting acidic peat. *Applied and Environmental Microbiology* **69**, 74-83 (2003).

26 Conrad, R. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiology Ecology* **28**, 193-202 (1999).
27 Keller, J. K. & Takagi, K. K. Solid-phase organic matter reduction regulates anaerobic decomposition in bog soil. *Ecosphere* **4**, 1-12 (2013).
28 Hodgkins, S. B. *et al.* Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences* **111**, 5819-5824 (2014).
29 Lipson, D. A., Jha, M., Raab, T. K. & Oechel, W. C. Reduction of iron (III) and humic substances plays a major role in anaerobic respiration in an Arctic peat soil. *Journal of Geophysical Research: Biogeosciences* **115** (2010).
30 Klüpfel, L., Piepenbrock, A., Kappler, A. & Sander, M. Humic substances as fully regenerable electron acceptors in recurrently anoxic environments. *Nature Geoscience* **7**, 195-200 (2014).
31 Christensen, T. R. *et al.* Thawing sub-arctic permafrost: Effects on vegetation and methane emissions. *Geophysical research letters* **31** (2004).
32 Whiticar, M. J. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. *Chemical Geology* **161**, 291-314 (1999).
33 McCalley, C. K. *et al.* Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478-481 (2014).
34 Mondav, R. *et al.* Discovery of a novel methanogen prevalent in thawing permafrost. *Nature communications* **5** (2014).
35 Stams, A. J. & Plugge, C. M. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology* **7**, 568-577 (2009).
36 Ishii, S. i., Kosaka, T., Hori, K., Hotta, Y. & Watanabe, K. Coaggregation facilitates interspecies hydrogen transfer between Pelotomaculum thermopropionicum and Methanothermobacter thermautotrophicus. *Applied and Environmental Microbiology* **71**, 7838-7845 (2005).
37 Wania, R. *et al.* Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP). *Geoscientific Model Development* **6**, 617-641 (2013).
38 Ji, M. *et al.* Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400, doi:10.1038/nature25014 (2017).
39 Welsh, D. T. Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS microbiology reviews* **24**, 263-290 (2000).
40 Rodrigues, D. F. *et al.* Architecture of thermal adaptation in an Exiguobacterium sibiricum strain isolated from 3 million year old permafrost: a genome and transcriptome approach. *BMC genomics* **9**, 547 (2008).
41 Maru, B., Bielen, A., Constanti, M., Medina, F. & Kengen, S. Glycerol fermentation to hydrogen by Thermotoga maritima: proposed pathway and bioenergetic considerations. *international journal of hydrogen energy* **38**, 5563-5572 (2013).

## Figure legends

**Figure 1. Genome-resolved view of the microbial communities at Stordalen Mire. a**, Schematic of the permafrost thaw gradient. Permafrost is light brown, active layer is blue (saturated peat) and dark brown (non-saturated peat). **b**, Community profile derived from MAG abundances (rows) across the active layer metagenomes (columns) from palsa (brown), bog (green) and fen (blue) samples. Black lines divide sites and depth ('E' denotes extra-deep). Red lines separate samples taken above (left) and below (right) the

water table. Numbers in brackets show total MAGs recovered and phylogenetic gain of Stordalen MAGs compared to publicly available genomes for each phylum. Red text indicates previously poorly represented phyla. **c**, Shannon diversity of each sample (filled circles) or averages for the sample's thaw stage and depth (empty circles).

**Figure 2. Carbon metabolism across the thaw gradient.** Boxplot headers and carbon compound boxes show degradation pathways. The large circles have outlines coloured by site, and contain smaller circles ('MAG abundances' coloured by phylum) representing the different MAGs encoding genes for the pathways shown. Circle size indicates MAG average relative abundance. The 'distribution boxplots' are coloured by site and stratified by depth. Boxplot y-axes indicate the cumulative relative abundances of the MAGs encoding the pathway of interest at the sites. Line thickness connecting the intermediates represents the average relative transcript 'expression' ('pathway expression') of pathway genes as transcripts per million reads mapped (TPM). Lines denote whether proteins were detected (solid) or not detected (dashed) in the metaproteomes. Relative abundances between sites were found to be significantly different for all pathways shown (see Methods). Coloured stars indicate relative abundances (ANOVA, p-value < 0.05) of pathways that are significantly different between depths.

**Figure 3. *Ca.* 'Acidiflorens' geochemical correlations and metabolic reconstruction**. **a**, Correlation of relative abundances of the *Ca.* 'Acidiflorens stordalenmirensis' with $\delta^{13}C$ of porewater $CH_4$ in bog sites. **b**, Correlation of the relative abundances of *Ca.* 'Methanoflorens stordalenmirensis' and *Ca.* 'Acidiflorens stordalenmirensis' in bog samples. **c**, Correlation of the relative abundances of *Ca.* 'Methanoflorens crillii' and *Ca.* 'Acidiflorens' sp. 2 in fen samples. **d**, Metabolic reconstruction of *Ca.* 'Acidiflorens'. Differential gene presence for the two lineages is indicated by colour-coding (Venn diagram). Functional units shown in grey are absent. Dotted lines indicate enzyme independent hydrogen movement. Purple cell cartoon shows *Ca.* 'Methanoflorens' spp. consuming hydrogen produced by *Ca.* 'Acidiflorens' spp.

# Methods

## Study Site

As described previously[34], Stordalen Mire is a peatland in northern Sweden (68 22' N, 19 03' E), 10 km southeast of Abisko. The three sub-habitats of the study site are common to northern wetlands, and together cover ~98% of Stordalen Mire's non-lake surface[10]. These proceed from well-drained palsas underlain by permafrost, and dominated by ericaceous and woody plants, to intermediate-thaw bogs with variable water table depth, dominated by *Sphagnum* spp. mosses, to fully thawed and inundated fens

dominated by sedges such as *Eriophorum angustifolium*. A thaw-associated shift in these habitats was documented between 1970 and 2000, as palsa collapsed and bogs and fens expanded by 3% and 54%, respectively[10]. These habitats exist in an intermingled mosaic, as is common in discontinuous permafrost zones, and the specific palsa, bog and fen that were sampled in this study are directly adjacent such that all cores were collected within a 120m radius.

This formation of wetlands after permafrost thaw is a widespread characteristic of peatlands affected by permafrost loss[42-45]. As frozen ground thaws it collapses, forming bogs and fens. Where this subsidence increases hydrologic connectivity, as at Stordalen, it can create a progression from ombrotrophic bogs to minerotrophic fens. A similar successional shift from bogs dominated by *Sphagnum* spp. to tall sedge fens has been observed in other northern peatlands[42,43,46,47]. The uncertainty surrounding the extent and characteristics of wetland formation from permafrost thaw is a critical limitation to modelling and understanding carbon–climate feedbacks[48,49]. Improved characterization of post-thaw microbial communities and carbon transformation processes, as advanced in this study, can directly address this uncertainty.

## Geochemistry

Across the thaw chronosequence, porewater $CH_4$ and $CO_2$ measurements and their $^{13}C$ isotopic composition were sampled as described previously[28]. The δ13C-$CH_4$ is affected by the δ13C-$CO_2$, because of the use or production of $CO_2$ during $CH_4$ generation, so the isotopic fractionation factor is used to report the isotopic separation of $CH_4$ and $CO_2$[50]. The $\alpha_C$ value reports the effective fractionation of C in $CH_4$, as the δ13C-$CH_4$ relative to source material represented by δ13C-$CO_2$. The effective fractionation factor of carbon in the porewater $CH_4$ relative to $CO_2$ ($\alpha_C$) was calculated as described previously [33,50].

$$\alpha_c = \frac{\delta 13C - CO_2 \ + 1000}{\delta 13C - CH_4 \ + 1000}$$

## DNA extraction and metagenome sequencing

DNA extractions were undertaken as described previously[34], with additional extractions from samples taken in 2012. Metagenome sequencing was performed for 2011 and 2012 using 100ng of the DNA in TruSeq Nano (Illumina) library preparation. For low concentration DNA samples, libraries were created using 1ng of DNA with the Nextera XT DNA Sample Preparation Kit (Illumina). 2012 libraries were sequenced on 1/12th of an Illumina HiSeq2000 lane producing 100bp paired-end reads, although some 2012 and 2011 samples were selected for deeper sequencing. Libraries from 2011 were sequenced 1/24th

of an Illumina NextSeq, producing 150bp paired end reads. See **Supplementary Data File 1** for details of sequencing depth per sample.

## Quantitative real time PCR

A quantitative polymerase chain reaction (qPCR) analysis was performed on selected samples to quantify microbial load. After pre-diluting 1/100, PCR was set up using 5µl of 2X SYBR Green/AmpliTaq Gold DNA Polymerase mix (Life Technologies, Applied Biosystems), 4µl of microbial template DNA and 1µl of primer mix. The 16S 1406F/1525R primer set (0.4µM) was designed to amplify bacterial and archaeal 16S rRNA genes: F - GYACWCACCGCCCGT and R - AAGGAGGTGWTCCARCC. The rpsL F/R primer set (0.2µM), used for inhibition control, amplifies Escherichia coli DH10B only: F - GTAAAGTATGCCGTGTTCGT and R - AGCCTGCTTACGGTCTTTA. Three dilutions 1/100, 1/500 and 1/1000 (microbial template DNA, 16S 1406F/1525R primer set) as well as an inhibition control (E. coli DH10B genomic DNA, rpsL primer set) were run in triplicate for each sample. The PCR was run on the ViiA7 platform (Applied Biosystems) including a cycle of 10 min at 95 °C (AmpliTaq activation) and 40 cycles of [15 s at 95 °C followed by 20 s at 55 °C and 30 s at 72 °C]. A melt curve was produced by running a cycle of 2 min at 95 °C and a last cycle of 15 s at 60 °C. The cycle threshold (Ct) values were recorded and analysed using ViiA7 v1.2 software.

CopyRighter[51] v0.46 was applied to qPCR counts to correct for 16S copy number variation. CopyRighter normalises the relative abundances across OTUs for each sample after dividing by the estimated copy number in a pre-computed table. The OTU genomic abundance is then obtained by multiplying by the total abundance number. A new CopyRighter database table was generated for the 2013 GreenGenes taxonomy (**Supplementary Data File 8**), with copy number estimates for leaf OTUs as the average copy number of IMG version 4.1 genomes mapped to GreenGenes genomes and clustered at 99% sequence identity, and for higher taxonomic levels inferred copy numbers for the clade common ancestor. The inferred copy numbers for higher taxonomies were propagated to descendent lineages without known copy numbers.

## SingleM

To determine microbiome diversity and community structure, SingleM was applied to reads from each sample (Woodcroft et al., unpublished, source code available at https://github.com/wwood/singlem).

## Diversity calculations

Shannon diversity[52] was calculated based on SingleM counts, rarefying to 100 sequences per marker gene when >100 sequences were detected and excluding samples otherwise. Vegan[53] was used to calculate the

diversity given the rarefied SingleM OTU table across each of the 15 marker genes, and the average was plotted in **Fig. 1**.

## Genome assembly and binning

Each sample's reads were assembled individually using CLC Genomics Workbench version 4.4 (CLC Genomics) with an estimated insert size of 50-500, generating 214 assemblies. Differential coverage binning was undertaken by mapping all reads from each sample of site (palsa, bog or fen) to all assemblies of that site, using BamM 'make' (Imelfort and Lamberton et al., unpublished, http://ecogenomics.github.io/BamM/) version 1.3.8-1.5.0, BWA 0.7.12[54], samtools[55], and GNU parallel[56]. Each sample's scaffolds were then binned using MetaBAT 3127e20aa4e7[57] using the sample's contigs and each of the BAM files as points of differential coverage.

The CheckM[58] v1.0.4 "lineage_wf" pipeline was used to determine completeness and contamination of the MAG bins through the identification and quantification of single copy marker genes, making use of pplacer 1.1 alpha 16[59]. Genomes estimated to be more than 70% complete and less than 10% contaminated were designated the "Stordalen MAGs".

## MAG dereplication and taxonomic classification

When calculating relative abundance, to alleviate multi-mapping issues, genomes were first dereplicated at 97% average nucleotide identity (ANI). First, amino acid identity was calculated between all genomes using the CompareM (v.0.0.17) AAI workflow ("comparem aai_wf", Parks, unpublished, https://github.com/dparks1134/CompareM). Genomes with an AAI of >95% were compared with each other using "calculate_ani.py" (Pritchard, unpublished, https://github.com/widdowquinn/scripts). Genomes with >97% ANI over >70% alignment were clustered together using single-linkage clustering, and the genome with highest quality in each cluster was chosen as the representative, where quality was calculated as "completeness - 4*contamination", as estimated by CheckM above. The cluster representative for each recovered MAG is provided in **Supplementary Data File 3**. The CompareM AAI workflow was also used to determine average amino acid identities between cluster representatives to determine the specific clades (**Supplementary Note 7**).

## Genome tree and phylogenetic inference of 1529 population bins

Phylogenetic inference was conducted in order to classify the MAG bins and used an in-house pipeline described in detail previously[60], the genome taxonomy database (GTDB v2.1.15) (Chaumeil & Parks, unpublished, https://github.com/Ecogenomics/GTDBNCBI). Briefly, a set of 122 archaeal and 120 bacterial specific single-copy marker genes were used to infer domain specific trees incorporating the 1529 MAGs, a reference set of genomes from NCBI (RefSeq[61] release 80), and the recently published

UBA genomes[62]. The concatenated alignment of these marker genes was created using HMMER v3.1.b2, and used as a basis for FastTree v2.1.9[63] tree building under the WAG + GAMMA model and using the approximately maximum likelihood method. This tree was then bootstrapped using genometreetk v0.0.35 (Parks, unpublished, https://github.com/dparks1134/GenomeTreeTk), calculating bootstrap support from 100 FastTree iterations. The associated taxonomy was derived using NCBI annotations, and was used to decorate the tree using tax2tree[64] and adjusted manually. Trees were visualised in ARB v6.0.6[65], and exported into ITOL[66] for further refinements before final editing in Inkscape. For the overall Bacteria and Archaea tree the dereplicated set of 647 genomes were selected in ARB and exported for viewing in ITOL. For the Acidobacteria tree (**Extended Data Fig. 1b**), Aminicenantes, including two Stordalen MAGs, and the recently reported *Ca*. Fischerbacteria[9] were included as likely classes within the Acidobacteria based on GTDB analysis (http://gtdb.ecogenomic.org/). The bootstrapped Newick trees for the overall Bacteria and Archaea trees are found in **Supplementary Data File 10 and 11**, using the alignments from **Supplementary Data Files 12 and 13.** Phylogenetic gain (**Fig. 1 & Extended Data Fig. 1**) was calculated using genometreetk pd_clade, and based on the added phylogenetic distances introduced to current phyla (comprising RefSeq release 80 and UBA genomes) by including the 1529 Stordalen MAGs.

## Calculation of relative abundance

To calculate the relative abundance of each genome in each lineage, reads from each sample were mapped to the set of dereplicated genomes using BamM “make”. Low quality mappings were removed with BamM v1.7.3 “filter” (minimum identity 95%, minimum aligned length 75% of each read) and the coverage of each contig calculated with BamM “parse” using “tpmean” mode, so calculating the coverage as the mean of the number of reads aligned to each position, after removing the highest 10% and lowest 10% of positions. The coverage of each MAG was calculated as the average of contig coverages, weighting each contig by its length in base pairs. The relative abundance of each lineage in each sample was calculated as its coverage divided by the total coverage of all genomes in the dereplicated set.

## Genomes which were differentially abundant by depth

To determine which lineages were differentially abundant between surface and deep samples, the set of relative abundances from each surface sample was compared to the set of relative abundances in the deep samples. The mean and statistical significance of the difference was calculated use R v3.3.2[67]. To determine the average amino acid identity between pairs of samples, the “aai_wf” of CompareM v0.0.7 (Parks, unpublished, https://github.com/dparks1134/CompareM) was used using the protein sequences predicted by Prokka as input.

## Annotation

Gene calling and preliminary annotation was undertaken with Prokka 1.11[68]. The genome was either annotated as Archaea or Bacteria, based in an inferred domain derived from the genome tree detailed above.

Annotation of glycoside hydrolase genes

All proteins predicted from all recovered genomes were screened using HMMSEARCH[69] using the dbCAN HMMs v5[70], using default parameters, then results post-processed to remove hits with e-value > 1e-18 and HMM coverage of < 0.35, where coverage was calculated as (hmm_to - hmm_from / qlen). Any genes with a hit passing these thresholds was then mapped to an EC number[71] using DIAMOND v0.8.27.89[72], with a database of all genes annotated with a fully defined (four number) E.C. number. This database of E.C. annotated genes was generated by gathering a list of GenBank identifiers of all characterised genes from each CAZy webpage[73] (listed http://www.cazy.org/Glycoside-Hydrolases.html) using a custom Ruby script and then downloading the corresponding protein sequences from GenBank.

Annotation of carbon metabolism

Annotation was undertaken using in-house scripts, which assign KEGG orthology to each gene via HMMs, taking the best hit and requiring an e-value of <1e-5. Encoding of whole pathways was inferred from genomes through the application of KEGG modules, both those available from KEGG as well as a number of custom modules (**Supplementary Data File 9**).

## Etymology

Description of '*Candidatus* Methanoflorens crillii' sp. nov.

'*Candidatus* Methanoflorens crillii' [´cril.lii. N. L. gen. n. '*crillii*', named after Prof. Patrick Crill, Stockholm University, Sweden, in recognition of his work on understanding of biogeochemical processes at the landscape scale (thawing permafrost) including greenhouse gases emission under the impact of climate change].

*Candidatus* Methanoflorens crillii sp. nov. is the second species recognised in the genus '*Candidatus* Methanoflorens'. The description is as provided by Mondav et al. (2014) for the genus with the following additional properties. The species can be differentiated from the recognised '*Ca*. M. stordalenmirensis' on the basis of phylogenetic analyses showing them to be monophyletic and sufficiently distinct average amino acid identity between encoded proteins.

Description of '*Candidatus* Acidiflorens stordalenmirensis' gen. nov., sp. nov.

‘*Candidatus* Acidiflorens stordalenmirensis’ [A.ci.di.flo´rens. N.L. n. acidum (from L. adj. acidus, sour), an acid; N.L. masc. substantive from L. masc. part. adj. florens, flourishing, to bloom; N. L. masc. n. Acidiflorens, an organism that blooms in acid; stor.da.len.mi.ren´sis. N.L. masc. adj. ‘stordalenmirensis’, of or belonging to Stordalen Mire, Sweden from where the species was characterised].

Description (brief). Phylogenetic analyses of genes/markers indicated that this species is different from all other recognised genera in the family Acidobacteriaceae.

Description of ‘*Candidatus* Changshengia’ gen. nov.

‘*Candidatus* Changshengia’ [Chan.gshen´gia. N. L. fem. n. ‘Candidatus Changshengia’, named in honor of Prof. Changsheng Li of The University of New Hampshire, a developer of the DeNitrification-DeComposition (DNDC) ecosystem model that contributed to our understating of the soil biogeochemical processes occurring in a variety of terrestrial ecosystems and climatic conditions. ]

*Candidatus* Changshengia gen. nov. is the second proposed and characterised genus in the phylum Dormibacteraeota. The delineation of genus is based on average amino acid identity between encoded proteins.

Description of ‘*Candidatus* Methanoflorentales’ ord. nov.

‘*Candidatus* Methanoflorentales’ (N.L. masc. adj. ‘*Candidatus* Methanoflorens’, type genus of the order; suff. –ales, ending to denote an order; N.L. fem. pl. n. ‘*Candidatus* Methanoflorentales’ the order of the genus ‘*Candidatus* Methanoflorens’).

The description is the same as given for the type genus ‘*Candidatus* Methanoflorens’ and the family ‘*Candidatus* Methanoflorentaceae’ Mondav et al. (2014) with the following modifications. The delineation of the order is determined by phylogenetic analyses showing that the Methanocellales would otherwise be paraphyletic. The order currently comprises two species ‘*Candidatus* M. stordalenmirensis’ and ‘*Candidatus* M. crillii’. The type genus is ‘*Candidatus* Methanoflorens’.

## Production and consumption rates of methane

Per-cell methane production and consumption rates were taken from studies of isolate cultures (for production 0.19 - 4.5 fmol $CH_4$ $cell^{-1}$ $h^{-1}$ [74-76] and for consumption 0.2 - 40 fmol $CH_4$ $cell^{-1}$ $h^{-1}$ [77-79]). Rates were taken as the mean of these rates for production and consumption rates, respectively.

## Bulk density measurements

Data is from one palsa core sample taken from July 2013. In the field, 50 $cm^3$ aliquots of fresh peat from each core section were removed and frozen. In the lab, each 50 $cm^3$ section of peat was weighed, freeze dried and then reweighed. Bulk densities were determined gravimetrically and calculated from the freeze-dried weights of the volumetric sections. Water contents were determined by the percent change in weight of the peat before and after freeze-drying.

## Metatranscriptomics

Metatranscriptome sequencing was conducted on select samples from 2010, 2011 and 2012, comprising four palsa, eight bog and twelve fen samples. ScriptSeq Complete (Bacterial) low-input library preparation kits (Epicentre, Madison WA) were used with 240 ng of sample RNA that had been co-extracted alongside the DNA from the initial sample material as input as described previously. Agilent 2100 Bioanalyzer and Agilent 2200 Tapestation (Agilent Technologies, Santa Clara CA) were used to check the quality of RNA and libraries during processing, with Qubit® (ThermoFisher Scientific, Waltham MA) used to determine quantity. These samples were run on 1/8th of a NextSeq (Illumina, San Diego CA) lane, with initial shallow runs conducted on 1/11th of a HiSeq (Illumina) and MiSeq (Illumina) lanes. Files originating from the same metatranscriptome libraries were concatenated before analysis.

SeqPrep (John, https://github.com/jstjohn/SeqPrep) was used to remove sequencing adaptors. PhiX contamination was removed by mapping the reads against the PhiX genome using BamM, and reads that aligned were removed. SortMeRNA v2.1[80] was used to remove non-coding RNA sequences (tRNA, tmRNA, 5S, 16S, 18S, 23S, 28S). To assign expression values to each gene, reads were first mapped in pairs to the dereplicated set of MAGs using BamM make, and filtered using BamM filter with cutoffs of 95% identity and 75% alignment. The count of reads mapped to each gene was calculated using DirSeq (https://github.com/wwood/dirseq, internally using bedtools[81]) based on any overlap of forward reads with the open reading frame of the gene, tabulating the sense and antisense mappings independently. To avoid the potential for DNA contamination of the RNA libraries to provide a misleading interpretation of a gene being expressed, the number of reads mapping in the sense direction were compared to the number mapping in the antisense direction using a one-sided binomial test. Genes with a significantly more reads mapping in the sense direction ($p<0.05$) were classified as ‘expressed’. For each significantly expressed gene, the number of antisense reads was subtracted from the number of sense reads to correct for metagenome contamination. These normalised expression estimates were used to calculate the TPM score[82], using only protein coding genes (CDS regions defined in the Prokka annotated GFF files) in each sample.

Pathway expression was calculated as the average expression of the steps within a pathway. If a pathway step included an enzyme complex, the average expression of each subunit was used as the expression value of that step. If a reaction could be catalysed by more than one enzyme, or if multiple copies of an enzyme were encoded by a genome, then their summed expression was used as the expression value of that step.

## Metaproteomics

Protein extraction, purification, and digestion:

Metaproteome analysis was conducted on 22 samples from 2012, collected from the same cores and depths as material used for metagenomes and metatranscriptomes. Three metaproteomes were created by pooling replicate cores (3x3 replicates). Sample nomenclature denotes year and month, followed by habitat (P = palsa, S = bog, E = fen), core number (123 indicates replicate cores 1, 2 and 3 were pooled) and depth (surface = S, mid = M, deep = D, extra deep = X). The 16 resulting metaproteomes were as follows: four palsa (20120600_P123M, 20120700_P3M, 20120700_P3D, 20120800_P2M) six bog (20120600_S123M, 20120600_S123D, 20120700_S2M, 20120700_S1D, 20120800_S1M, 20120800_S1X), and six fen samples (20120700_E3M, 20120700_E3D, 20120700_E2X, 20120800_E2M, 20120800_E2D, 20120800_E3D) (**Supplementary Data File 1**). Proteins were extracted and digested using substantial modifications of methods developed previously for our site[34]. Briefly, samples were thawed and 35 g of peat per sample was split equally into two 50 ml tubes, and sodium dodecyl sulphate (SDS)-resuspension buffer was added to a final volume of 30 ml. SDS-resuspension buffer (pH 8) was freshly prepared as: (1) an SDS buffer of 50 mM dithiothreitol (DTT) in 10 ml of 4% SDS, (2) a separate resuspension buffer of 50 mM trisaminomethane (Tris) Buffer (2.21 g Trizma-HCl and 4.36 g Trizma Base (Millipore Sigma, St. Louis, MO, USA)), 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and HPLC-grade water up to 1 L, (3) the 10 ml SDS buffer (warmed at 60 °C for 2 min) was mixed with 40 ml of resuspension buffer, and the final pH was adjusted to 8. Samples were vortexed for 10 min using a tabletop vortexer with adapters for 50 ml conical tubes, and then 10 g of 0.1 mM glass beads (Qiagen, Hilden, Germany) were added, followed by 30 min of vortexing. Samples were centrifuged at 3,000 x g for 20 min, the supernatant transferred to a new tube and centrifuged at 4,800 g for 20 min. The supernatant was transferred to a new tube, to which 100% trichloroacetic acid (TCA) was added to a final concentration of 30%. Samples were shaken and then stored at 4 °C overnight.

Samples were centrifuged at 4,800 x g for 1 hr 30 min at 4 °C, and then supernatant was decanted and pellets from the same sample were combined. The following steps were repeated three times: pellets were washed with 1 ml cold acetone, placed on ice for 5 min, vortexed briefly, centrifuged at 24,000 x g for 25

min at 4 °C, and then supernatant was removed. Pellets were dried under $N_2$ gas, and then 1-1.5 ml of denaturing buffer was added. Denaturing buffer was prepared as follows: (1) a digestion buffer was prepared with 4.88 g of Trizma-HCl, 2.30 g of Trizma Base, and 1.11 g of $CaCl_2$ brought to a 1L volume with HPLC-grade water, (2) Guanidine-HCl was added to digestion buffer to a final concentration of 6M in a 50-ml tube. Samples were incubated at 60 °C for one hour, vortexing for 5 seconds every 10 min, then transferred to a new tube, to which digestion buffer was added to a final volume of 15 ml. Proteins were digested by adding 20 µg trypsin (NEB) and incubating on a nutating mixer at 37 °C overnight.

A further 10 µg of trypsin was added to each sample, followed by incubation on a nutating mixer at 37 °C for 4 hr. DTT was added (approximately 100 mg), then samples were returned to the 37 °C nutating mixer for 45 min and then centrifuged at 3,000 rpm for 5 min at 20 °C. The supernatant was transferred to a new tube. Similar to previously reported peptide purification protocols[83], Sep-Pak Plus C18 cartridges (Waters Corporation, Milford, MA, USA) were conditioned with 10 ml of acetonitrile + 0.1 % formic acid, then washed with 10 ml of 0.1% formic acid (in water), then the samples were added and the flow-through was refiltered through the cartridges three additional times, followed by a wash with 10 ml of 0.1% formic acid (in water). Peptides were eluted with 5 ml of acetonitrile + 0.1 % formic acid, then added to 0.45 µm Ultrafree-MC filter tubes according to the manufacturer's protocol (Millipore Sigma).

<u>Peptide fractionation and mass spectroscopy:</u>

A bicinchoninic acid assay (Thermo Scientific, Rockford, IL) was performed to determine the peptide mass in each sample. The samples were then diluted with 10 mM ammonium formate, pH 10 ('buffer A') to a volume of 930 µl, centrifuged at 10k x g for 2-5 minutes to remove any precipitates, and transferred to snap-cap ALS vials. The diluted samples (pH 10) were resolved on a XBridge C18, 250x4.6 mm, 5 µM with 4.6x20 mm guard column (Waters, Milford, MA). Separations were performed at 0.5 ml/min using an Agilent 1100 series HPLC system (Agilent Technologies, Santa Clara, CA) with mobile phases (A) buffer A and (B) buffer A/acetonitrile (10:90). The gradient was adjusted from 100% A to 95% A over the first 10 min, 95% A to 65% A over minutes 10 to 70, 65% A to 30% A over minutes 70 to 85, maintained at 30% A over minutes 85 to 95, re-equilibrated with 100% A over minutes 95 to 105, and held at 100% A until minute 120. Fractions were collected every 1.25 minutes (96 fractions over the entire gradient) and every 12th fraction were pooled for a total of 12 fractions per sample. All fractions were dried under vacuum and 20 µL of nanopure water was added to each fraction for storage at -20°C until LC-MS/MS analysis.

Fractions were analyzed by reversed-phase LC-MS/MS using a Waters nanoAquity™ UPLC system (Millford, MA) coupled with a Q-Exactive Plus hybrid quadrupole/Orbitrap mass spectrometer from Thermo Fisher Scientific (San Jose, CA). The analytical column was packed in-house by slurry packing 3-µm Jupiter $C_{18}$ stationary phase (Phenomenex, Torrence, CA) into a 70-cm long, 360 µm OD x 75 µm ID fused silica capillary tubing (Polymicro Technologies Inc., Phoenix, AZ). Mobile phases consisted of 0.1% formic acid in water (MP-A) and 0.1% formic acid in acetonitrile (MP-B). Samples were adjusted to a concentration of ~ 0.1 µg/µL and 5 µL injections were directly loaded onto the analytical column at a flow rate of 300 nL/min and 1% MP-B. The full loading, gradient elution, and column regeneration profile was as follows (min:%MP-B); 0:1, 30:1, 32:8, 50:12, 105:30, 110:45, 120:90, 125:90, 130:1, 170:1. Data acquisition (100 min) was started at the end of the sample loading period (30 min). The analytical column was coupled to the Q-Exactive using a home-built nanospray adapter interface with 2.2 kV applied to achieve electrospray ionization. The MS inlet was maintained at a temperature of 300ºC. A precursor scan was performed from m/z 300 to 1800 at a resolution of 30k and an automatic gain control (AGC) of 3e6. Operated in data dependent mode, the top 12 most intense ions from the precursor scan were selected for high energy collision dissociation (HCD) MS/MS at a resolution of 17.5k, AGC of 1e5, isolation window of 2 m/z, and a max ion time of 100 ms. Only ions identified as having a +2 charge or higher were subjected to HCD and subsequently excluded from further analysis for 30 sec thereby allowing for deeper coverage. In total, 192 mass spectra (MS) files were generated (12 fractions for each of the 16 samples).

<u>Database search and expression analysis:</u>

A sensitive and universal database search tool, MSGFPlus[84] v2017.01.13, was used to conduct the metaproteome searches in this study. Prior to searching the metaproteomes, the mass spectrometer RAW output files were converted to the mzML format using msConvert of ProteoWizard[85] 3.0.10200, accepting the default parameters. The mzML files were then searched against a targeted protein database containing protein sequences predicted from the metagenome-assembled genomes across the permafrost thaw gradient and involved in metabolic pathways examined in the manuscript, as well as the entire CDS regions of *Ca.* 'Acidiflorens', *Ca.* 'Methanoflorens' and AD3 (**Supplementary Data File 14**). Proteins in the targeted database were dereplicated at 100% amino acid identity using usearch v9.2.64 (--fastx_uniques)[86] after converting all isoleucine residues to leucine (due to identical masses). In order to calculate the false discovery rate (FDR), a parallel search of a decoy protein database was conducted by using the (-tda 1) parameter of MSGFPlus during the indexing and searching steps of the targeted database. After conducting the searches, the FDR was calculated as: FDR(t) = #DecoyPSMs at (t) / #TargetPSMs at (t), where (t) is the highest Q-value that gives a FDR of ≤ 1%, and #DecoyPSMs and

#TargetPSMs are the numbers of Decoy and Target Peptide-Spectrum Matches (PSM), respectively, at that Q-value threshold. Only PSMs with Q-values of ≤ (t = 0.0145) were considered in the results of this study. The maximum precursor mass tolerance allowed during the searches was specified by the (-t 20 ppm) parameter for parent mass tolerance and the (-ti "-1,2") parameter for isotope error range. Trypsin was specified as the digestion method (-e 1) and only full tryptic digestion was allowed (-ntt 2). Minimum and maximum peptide lengths to consider were 6 and 50, respectively. Minimum and maximum precursor charges to consider were 2 and 5, respectively. For each spectral scan, only the PSMs with the highest MSGF score (-n 1) were considered for subsequent analyses.

The detection of a unique peptide in at least one sample was considered evidence of expression of a specific enzyme. A specific pathway was designated as expressed when at least one enzyme from a pathway was detected in the metaproteome. Further, that enzyme had to be encoded in a genome where >50% of genomes in the 97% ANI genome cluster encoded all steps of that pathway. Only a very small number of proteins (<1%) were detected where the sequence of that protein was 100% identical to a protein from a different ANI genome cluster.

The limited number of proteins detected for the fen hydrolysers is likely a consequence of the diversity and complexity of the fen community, and the variety of proteins produced, compared to the palsa and bog. The metaproteomes were searched against a targeted protein database containing only those sequences predicted from the MAGs. Peptide-spectrum matching requires 100% matches, which means that variations in sequence due to population heterogeneity/diversity, which is greatest in the fen (**Fig. 1; Extended Data Fig. 2**), are not captured, reducing the number of successful matches. Our approach is conservative following the genome-centric focus, and used purely to show that the populations of interest are translating proteins in situ.

## Statistical analyses

Analysis of variance (ANOVA) tests, Mann-Whitney U tests, and least squares regressions were calculated using R[67]. Boxplots were created using ggplot2[87], with the boxes representing the 1st to 3rd quartiles and the whiskers the highest (upper whiskers) or lowest (lower whiskers) observation within 1.5x the interquartile range. The centre line represents the mean. **Fig. 2** sample numbers were as follows: palsa surface n = 18, palsa mid n= 18, palsa deep n = 17, bog surface n = 19, bog mid n = 20, bog deep n = 20, bog extra deep n = 6, fen surface n = 23, fen mid n = 18, fen deep n = 23, fen extra deep n = 6 biologically independent samples. A square root scale was used on the Y axis. P-values for significant differences amongst depths assessed with ANOVA were $4\times10^{-7}$, $2\times10^{-5}$, and $2\times10^{-5}$ for palsa, bog and fen respectively for cellulose degradation, and $4\times10^{-2}$, $3\times10^{-8}$ and n/a for xylan degradation, $5\times10^{-3}$, n/a, n/a

for xylose degradation (dehydratase), n/a, $2\times10^{-12}$, $1\times10^{-4}$ for xylose degradation (oxidoreductase), $3\times10^{-4}$, $9\times10^{-7}$, $6\times10^{-3}$ for xylose degradation (isomerase), $2\times10^{-4}$, $5\times10^{-9}$, n/a for lactate fermentation, n/a, $3\times10^{-2}$, n/a for ethanol fermentation, $8\times10^{-5}$, $5\times10^{-7}$, n/a for propionate fermentation, n/a, $2\times10^{-4}$, $5\times10^{-2}$ for hydrogenotrophic methanogenesis, n/a, $2\times10^{-3}$, $4\times10^{-3}$ for acetoclastic methanogenesis. **Fig. 3** sample numbers were as follows: panel **a**, n = 47, panel **b**, n = 65, panel **c**, n = 70 biologically independent samples.

To test for associations between microbial populations and geochemical variables, the abundance of genus- to family- level lineages was calculated using phylogenetic tree insertion of open reading frames derived from individual reads into single-copy ribosomal protein marker trees using GraftM[88] and phylogenetic trees derived from both contigs assembled from data presented here and reference datasets (data not shown). To avoid statistical complications arising from multiple hypothesis testing, only the ten most abundant lineages in the bog and fen were tested for significance using least squares regression. Lineages correlating significantly were then linked to MAGs for correlations between individual MAGs and geochemical variables as well as between *Ca*. 'Acidiflorens' sub-lineages.

## Code availability

The above methods indicate the source of the code and programs used for analyses within the relevant sections.

## Figure generation

Manuscript figures were generated using custom R[67] scripts, ggplot2[87], spacemacs (http://spacemacs.org/), Rstudio[89], arb[65], d3js (https://d3js.org/), Inkscape (https://inkscape.org/) and Adobe Illustrator (http://www.adobe.com/au/products/illustrator.html).

## Data Availability

Data described in this manuscript are submitted under NCBI BioProject accession number PRJNA386568. MAGs were deposited at DDBJ/ENA/GenBank under the accession numbers provided in Supplementary Data File 3, and the initial versions are described in this paper. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[90] partner repository with the dataset identifier PXD009096 and 10.6019/PXD009096. Supplementary Data Files 1 - 9 are available with the online version of this manuscript. Supplementary Data Files 10 - 15 are available on figshare DOI: 10.6084/m9.figshare.6233660.

## Extended Data Legends

**Extended Data Figure 1**. **Phylogenetic distribution of MAGs recovered from Stordalen Mire**. **a**, Phylogenetic tree of 647 dereplicated MAGs. Numbers in brackets show total MAGs recovered and phylogenetic gain of Stordalen MAGs compared to publicly available genomes for each phylum. Red text indicates previously poorly represented phyla. **b**, Acidobacteria subtree showing the *Ca*. 'Acidiflorens' lineages. **c**, Eremiobacteraeota subtree incorporating the CARN1[91] MAG. **d**, Dormibacteraeota subtree, showing *Ca*. 'Changshengia'. **e**, Subtree of *Ca*. 'Methanoflorentales' MAGs, and closest neighbouring orders. In panels **b** – **e**, pie charts show phylogenetic gain, red lines indicate Stordalen MAGs, black lines indicate public genomes, blue triangles indicate clustered public genomes, red triangles indicate clustered Stordalen MAGs. Black dots indicate bootstrap values 70-100%.

**Extended Data Figure 2**. Microbial community profile of the thaw gradient. In **a**, the relative abundance of each phylum as estimated through the recovery of 16S rRNA gene reads is shown, averaged within each thaw stage. The 15 phyla with the highest relative abundance across all samples are shown. In **b**, the number of MAGs recovered from each of these phyla is plotted, showing that broadly, MAGs recovered are from lineages highest in abundance. In **c**, a principal coordinates analysis of weighted UniFrac compositional differences between samples is shown, based upon average coverage of each recovered genome of reads mapped to the dereplicated genome set. Colours indicate thaw stage: brown = palsa (P), green = sphagnum/bog (S), blue = eriophorum/fen (E). Depth: S = surface, M = mid-depth, D = deep, X = extra-deep. Goodness of fit was 0.57 for PCoA 1 and 0.65 for PCoA 2. Sample numbers: n = 53, 65 and 70 biologically independent samples for palsa, bog and fen respectively. In **d**, quantitative PCR analysis of samples taken in 2012 is shown. The number of cells per gram of soil is shown for three depths at the three thaw stages, after correcting for 16S rRNA gene copy number variation (see **Methods**). Fen samples contained significantly more cells per gram of soil than bog and palsa samples (average 2.6×, p-value 7e-8, n=103, two sided Mann-Whitney U test). Sample numbers: n = 8, 9, and 8 for biologically independent samples palsa surface, mid and deep respectively, n = 9, 8, 9 and 10, 7 and 9 for bog and fen respectively. Panels **e** and **f** show relative abundances of phyla and classes within the Proteobacteria across the thaw gradient, respectively. The depth of each sample is indicated by the colour of the box (surface: red, mid-depth: green, deep: blue, extra-deep purple). Each data point is the sum of relative abundances of all lineages assigned to the phylum in a sample after adding a 0.1% pseudocount to all phyla (so the y-axis is not dominated by small values visually). Boxplots are shown plotted on a log-scale y-axis, with phyla and classes ordered by decreasing average relative abundance across all samples. Relative abundance was calculated based on the fraction of the community with recovered genomes (see **Methods**). Sample numbers: n = 53, 65 and 70 biologically independent samples for palsa, bog and fen respectively.

**Extended Data Figure 3.** Prevalence of individual MAGs across the thaw gradient. In **a**, the number of samples where each Stordalen MAG is present at >1% relative abundance is shown among each stage of the thaw gradient. Vertical red lines indicate the number of samples sequenced in total from that environment. Only one MAG "Deltaproteobacteria_fen_1087" was found in a high abundance across fen sites, detected at >1% relative abundance in 96% of fen sites. In **b**, the same information is shown stratified by depth of the sample in the soil column. The specific MAGs prevalent are detailed in **Extended Data Table 1**, showing that a small number of populations were prevalent at a specific depth of a specific site. Panel **c** shows Stordalen genomes that significantly changed in abundance with depth. For each site, genomes which show the largest absolute difference in abundance between shallow and deep samples are shown. Genomes which are more abundant in shallow samples compared to deep are positive, and those more abundant in deep samples relative to shallow samples are negative. Only those lineages with a mean absolute difference of >1% and are significantly different (p-value < 0.05, two sided Mann-Whitney U test) are shown. Sample numbers: n = 53, 65 and 70 biologically independent samples for palsa, bog and fen respectively. Each bar indicates a 97% dereplicated MAG which changes in relative abundance between surface and deep samples and the colour of each indicates the phylum the genome belongs to. The fen is less stratified between the surface and deep, which is reflected in the fewer population abundances significantly changing in abundance between shallow and deep samples. Recovered congeneric genomes which showed significant but inverse differential abundance between surface and deep samples are shown in **Supplementary Data File 7**. Genomes depicted in c in order are Acidobacteria_palsa_348 = 1, Acidobacteria_palsa_246 = 2, Actinobacteria_palsa_463 = 3, Actinobacteria_palsa_558 = 4, Acidobacteria_palsa_312 = 5, Alphaproteobacteria_palsa_929 = 6, Actinobacteria_palsa_504 = 7, Acidobacteria_palsa_125 = 8, WPS2_palsa_1515 = 9, Acidobacteria_palsa_289 = 10, WPS2_palsa_1516 = 11, Acidobacteria_palsa_310 = 12, Alphaproteobacteria_palsa_913 = 13, Actinobacteria_palsa_693 = 14, Actinobacteria_palsa_465 = 15, Actinobacteria_palsa_691 = 16, Alphaproteobacteria_palsa_895 = 17, Actinobacteria_palsa_505 = 18, Actinobacteria_bog_593 = 19, Actinobacteria_palsa_462 = 20, Acidobacteria_palsa_199 = 21, Acidobacteria_palsa_362 = 22, Acidobacteria_palsa_313 = 23, Gammaproteobacteria_palsa_1209 = 24, Acidobacteria_palsa_267 = 25, Planctomycetes_palsa_1347 = 26, Acidobacteria_palsa_143 = 27, Verrucomicrobia_palsa_1397 = 28, Actinobacteria_palsa_641 = 29, Actinobacteria_palsa_733 = 30, Acidobacteria_palsa_420 = 31, Actinobacteria_palsa_736 = 32, Verrucomicrobia_palsa_1413 = 33, Alphaproteobacteria_palsa_910 = 34, Acidobacteria_palsa_286 = 35, Acidobacteria_palsa_122 = 36, Acidobacteria_palsa_343 = 37, Deltaproteobacteria_palsa_1114 = 38, Gemmatimonadetes_palsa_1248 = 39, Acidobacteria_palsa_340 = 40, Acidobacteria_palsa_141 = 41, Alphaproteobacteria_palsa_922 = 42, WPS2_palsa_1496 = 43, Actinobacteria_bog_635 = 44, Actinobacteria_bog_766 = 45,

Actinobacteria_bog_592 = 46, Gammaproteobacteria_bog_1200 = 47, Actinobacteria_bog_594 = 48, Acidobacteria_bog_329 = 49, Verrucomicrobia_bog_1475 = 50, Actinobacteria_bog_723 = 51, Acidobacteria_bog_233 = 52, Verrucomicrobia_bog_1402 = 53, WPS2_bog_1492 = 54, Alphaproteobacteria_bog_899 = 55, WPS2_bog_1527 = 56, Actinobacteria_bog_769 = 57, Acidobacteria_bog_377 = 58, Actinobacteria_bog_637 = 59, FCPU426_bog_1183 = 60, Alphaproteobacteria_bog_900 = 61, Acidobacteria_bog_234 = 62, WPS2_bog_1502 = 63, Verrucomicrobia_bog_1421 = 64, Gammaproteobacteria_bog_1206 = 65, Alphaproteobacteria_bog_908 = 66, Betaproteobacteria_bog_994 = 67, Acidobacteria_fen_416 = 68, Actinobacteria_fen_548 = 69, Acidobacteria_bog_445 = 70, Acidobacteria_bog_96 = 71, Acidobacteria_bog_202 = 72, Actinobacteria_fen_455 = 73, AD3_bog_854 = 74, Acidobacteria_bog_218 = 75, Actinobacteria_bog_806 = 76, Acidobacteria_bog_390 = 77, Actinobacteria_bog_524 = 78, Euryarchaeota_bog_81 = 79, Verrucomicrobia_bog_1459 = 80, AD3_bog_876 = 81, Actinobacteria_bog_808 = 82, Acidobacteria_bog_226 = 83, Actinobacteria_bog_576 = 84, Acidobacteria_bog_406 = 85, Acidobacteria_fen_408 = 86, Deltaproteobacteria_fen_1088 = 87, Nitrospirae_fen_1304 = 88, Bacteroidetes_fen_982 = 89, Bacteroidetes_fen_956 = 90, Acidobacteria_fen_335 = 91, Euryarchaeota_fen_63 = 92, Gammaproteobacteria_fen_1191 = 93, Deltaproteobacteria_fen_1087 = 94, Gammaproteobacteria_fen_1218 = 95, Gammaproteobacteria_fen_1219 = 96, Actinobacteria_fen_730 = 97, Deltaproteobacteria_fen_1138 = 98, Chloroflexi_fen_1050 = 99, Actinobacteria_fen_453 = 100, Acidobacteria_fen_408 = 101, Actinobacteria_fen_548 = 102, Chloroflexi_fen_1019 = 103, Acidobacteria_fen_414 = 104, Actinobacteria_fen_455 = 105.

**Extended Data Figure 4**. Cellulase (**a**, **b**), xylanase (**c**, **d**) and β-glucosidase (**e**, **f**) gene expression across the thaw gradient. Samples analysed with metatranscriptomics are described by the date of sampling, core number and depth. In **a**, **c** and **e**, the relative contribution of each phylum to the total TPM (transcripts per million reads mapped) of the enzyme class observed in the metatranscriptomes is shown. In **b**, **d** and **f**, the total TPM of all expressed genes in the sample is charted.

**Extended Data Figure 5.** Monosaccharide degradation pathway prevalence at Stordalen Mire. In **a**, as in **Fig. 2**, 97% dereplicated MAGs are shown as circles ('MAG abundance'), where the radius of the circle represents the average relative abundance of that genome in the palsa, bog or fen. In **b**, as in **Fig. 2**, the total relative abundance of genomes encoding the pathway is shown amongst the entire community. Sample numbers: n = 53, 65 and 70 biologically independent samples for palsa, bog and fen respectively.

**Extended Data Figure 6.** Xylose degradation pathways at Stordalen Mire. **b**, Venn diagram showing how each xylose breakdown pathway is shared among the Stordalen Mire MAGs. Percentages represent the proportion compared to all Stordalen genomes encoding a xylose degradation pathway. In the metaproteomes, genomes Acidobacteria_bog_390, Actinobacteria_fen_455 and Actinobacteria_bog_808 expressed a protein specific to oxidoreductase pathways and a protein specific to the isomerase pathway. In the metatrascriptomes, Acidobacteria_palsa_248, Acidobacteria_bog_370, Acidobacteria_bog_390, Actinobacteria_fen_455, Actinobacteria_bog_586, Actinobacteria_bog_808 and Planctomycetes_fen_1346 expressed a protein specific to oxidoreductase pathways and a protein specific to the isomerase pathway. **c**-**h**, Gene expression of xylose degradation pathways. Average expression of genes in the canonical bacterial xylose isomerase (**c**, **d**), oxidoreductase (**e**, **f**) and xylanate dehydratase pathways (**g**, **h**) are depicted across the thaw gradient. Samples analysed with metatranscriptomics are described by the date of sampling, core number and depth. In **c**, **e** and **g**, the relative contribution of each phylum to the total TPM (transcripts per million reads mapped) of the enzyme class observed in the metatranscriptomes is shown. In **d**, **f** and **h**, the total TPM of all expressed genes in the sample is charted.

**Extended Data Figure 7**. Gene expression of fermentation pathways. Samples analysed with metatranscriptomics are described by the date of sampling, core number and depth. In **a**, **c**, **e** and **g**, the total TPM (transcripts per million reads mapped) of each fermentation pathway in the metatranscriptomes is charted. In **b**, **d**, **f** and **h**, the relative contribution of each phylum to the total TPM of each pathway is shown.

**Extended Data Figure 8**. **a**, $CO_2$ and $CH_4$ concentrations in porewater derived from the bog and fen. The blue line shown is a line of best fit, forced through the origin. Dots indicate the samples, with colors indicating the sample depth. The concentrations are correlated, and the $CH_4$ concentrations are much less than the $CO_2$ concentrations in both sites. Sample numbers: n = 51 (bog) and 61 (fen) biologically independent samples. **b**, Methanogenesis versus methanotrophy rates. Each point represents the average relative abundance of methanotrophs and methanogens across all samples in a single core, multiplied by the rate of methane generation or consumption inferred from previous culture-based measurements (2.345 and 20.1 fmol $CH_4$ / hr / cell of methanogenesis and methanotrophy, respectively, see **Methods**). The line represents the 1:1 ratio. Inferred fluxes were calculated using relative abundance of methanogenic or methanotrophic lineages so rates are only intended for comparison between the x and y axes, rather than as an absolute measure of $CH_4$ flux. Methanotrophy appears to mitigate a significant proportion of the $CH_4$ generated in the bog sites. **c**, Correlation of the relative abundance of *Ca.* `Methanoflorens stordalenmirensis' with the isotopic fractionation of methane ($\alpha_C$) dissolved in paired porewater samples taken from the bog is shown. Previously observed in 2011 using 16S rRNA gene amplicon sequencing[33],

the correlation is confirmed here using genome-centric metagenomic techniques on the 2011 samples, as well as in a new year of sampling in 2012. Sample numbers: n = 23 (2011) and 24 (2012) biologically independent samples. **d** and **e**, Expression of methanogenesis marker gene *mcrA* across the thaw gradient is shown. Samples analysed with metatranscriptomics are described by the date of sampling, core number and depth. In **d**, the relative contribution of each methanogenic order to the total TPM (transcripts per million reads mapped), **e**, of all *mcrA* genes in the metatranscriptome is shown. Metaproteomes revealed the expression of 289 hydrogenotrophic McrA proteins across 13 samples, as well as 78 acetoclastic McrA proteins across eight samples (**Supplementary Data File 2**). **f**, linear regression analysis for predicting effective fractionation ($\alpha_c$) of $CH_4$ from environmental variables and *Ca*. 'Methanoflorens stordalenmirensis' abundances in the bog are tabulated. *Ca*. 'Methanoflorens stordalenmirensis' abundance exceeds bulk geochemical parameters in predicting the effective fractionation of $CH_4$. Each line is the result of a linear regression of the specified measurement against the $\alpha_c$ of $CH_4$ in bog porewater samples taken in 2011 and 2012 (n = 47 biologically independent samples).

**Extended Data Figure 9**. Candidate phylum Dormibacteraeota (AD3) genus *Ca*. 'Changshengia' at Stordalen Mire. **a**, Total relative abundance of the genus *Ca.* 'Changshengia' correlated with the fraction of the concentration of C mineralised to $CO_2$ versus $CH_4$ in the bog porewater samples ($R^2$ 0.19, p-value = 0.001, n = 51 biologically independent samples). Each point represents an individual sample from 2012, with its colour representing the depth in the core that the sample was taken from. **b**, Metabolic reconstruction of genomes belonging to the candidate phylum AD3 genus *Ca*. 'Changshengia' correlating with the $CH_4$:$CO_2$ concentration ratio in porewater from 2012 bog samples. Genomes from four clades within the AD3 were assembled from across Stordalen Mire. Enzyme colour indicates the families that share that metabolic potential, as outlined in the legend on the left. Arrow colouring indicates whether expression was detected (red arrows) or not detected (black arrows) for genes encoding the enzyme in any of the 24 metatranscriptomes. Orange stars indicate detection of protein expression in any of the 22 metaproteomes from the *Ca.* 'Changshengia' and related genomes. All four lineages encode the potential to oxidise glycerol anaerobically through glycerol transporter (*glpF*), glycerol kinase (*glpK*) and a membrane bound glycerol-3-phosphate dehydrogenase (*glpABC*), entering glycolysis via dihydroxyacetone phosphate processed to glyceraldehyde-3-phosphate by the triosephosphate isomerase (*tpiA*). Other glycerol derivatives such as glycerol-3-phosphate could be imported (*glpT*) by this and other family members, and dihydroxyacetone phosphate can also be processed using the PTS-dependent dihydroxyacetone kinase (*dhaLMK*) complex. Sinks for the electrons generated from the oxidation of glycerol also varied between the different lineages, with *Ca*. 'Changshengia' and clade 1 having a H+ translocating complex I NADH:oxidoreductase, while clade 1 also has a high affinity cytochrome oxidase

complex IV, clade 2 genomes only encode a nitrate reductase (*narGHI*) and clade 4 genomes only a fumarate reductase (*sdhABCD*). These differences likely lead to the differentiation of the niches that each lineage occupies across different sites and depths of the mire. Lineages were considered positive for genes or complexes based on the presence of sequences with 80% homology in 50% of the genomes. **c,** Phylogenetic subtree showing the family groupings of AD3 for the metabolic analysis. Representative genomes from the 97%ANI dereplication are indicated in red. Bootstrap support is indicated at the nodes for values over 70% or 90% in grey and black, respectively. Blue clade indicates cluster of seven UBA and RefSeq genomes.

**Extended Data Table 1.** Genomes with high prevalence in specific sites and depths. Genomes shown are representative genomes from the 97% dereplicated set which are present at >1% relative abundance in >90% of samples from the site and depth shown in the "Environment" column. The "Num present" column indicates the number of samples it is found in and "Total samples" indicates the total number of samples available for that environment.

**Extended Data Table 2.** Overview of proteins detected with metaproteomics. The third column shows the ten most abundant KEGG Orthology (KO) groups detected, where the total spectral count for that group was two or more. The fourth column shows the relative abundance of spectral counts from each phylum.


9 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7** (2016).

10 Johansson, T. *et al.* Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Global Change Biology* **12**, 2352-2369 (2006).

28 Hodgkins, S. B. *et al.* Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proceedings of the National Academy of Sciences* **111**, 5819-5824 (2014).

33 McCalley, C. K. *et al.* Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478-481 (2014).

34 Mondav, R. *et al.* Discovery of a novel methanogen prevalent in thawing permafrost. *Nature communications* **5** (2014).

42 Vitt, D. H., Halsey, L. A. & Zoltai, S. C. The changing landscape of Canada's western boreal forest: the current dynamics of permafrost. *Canadian Journal of Forest Research* **30**, 283-287 (2000).

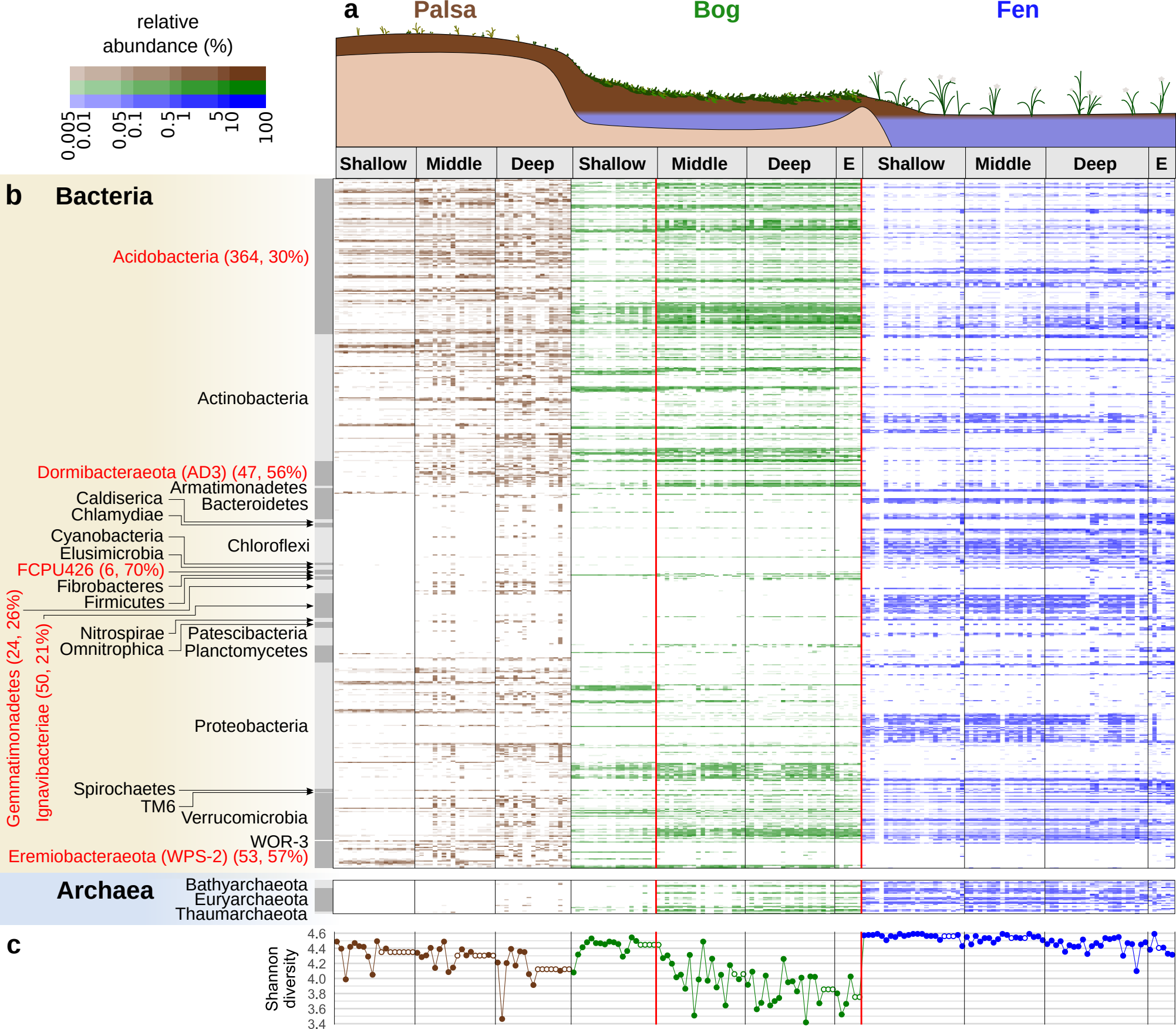
43 Jorgenson, M. T., Racine, C. H., Walters, J. C. & Osterkamp, T. E. Permafrost degradation and ecological changes associated with a warmingclimate in central Alaska. *Climatic Change* **48**, 551-579 (2001).

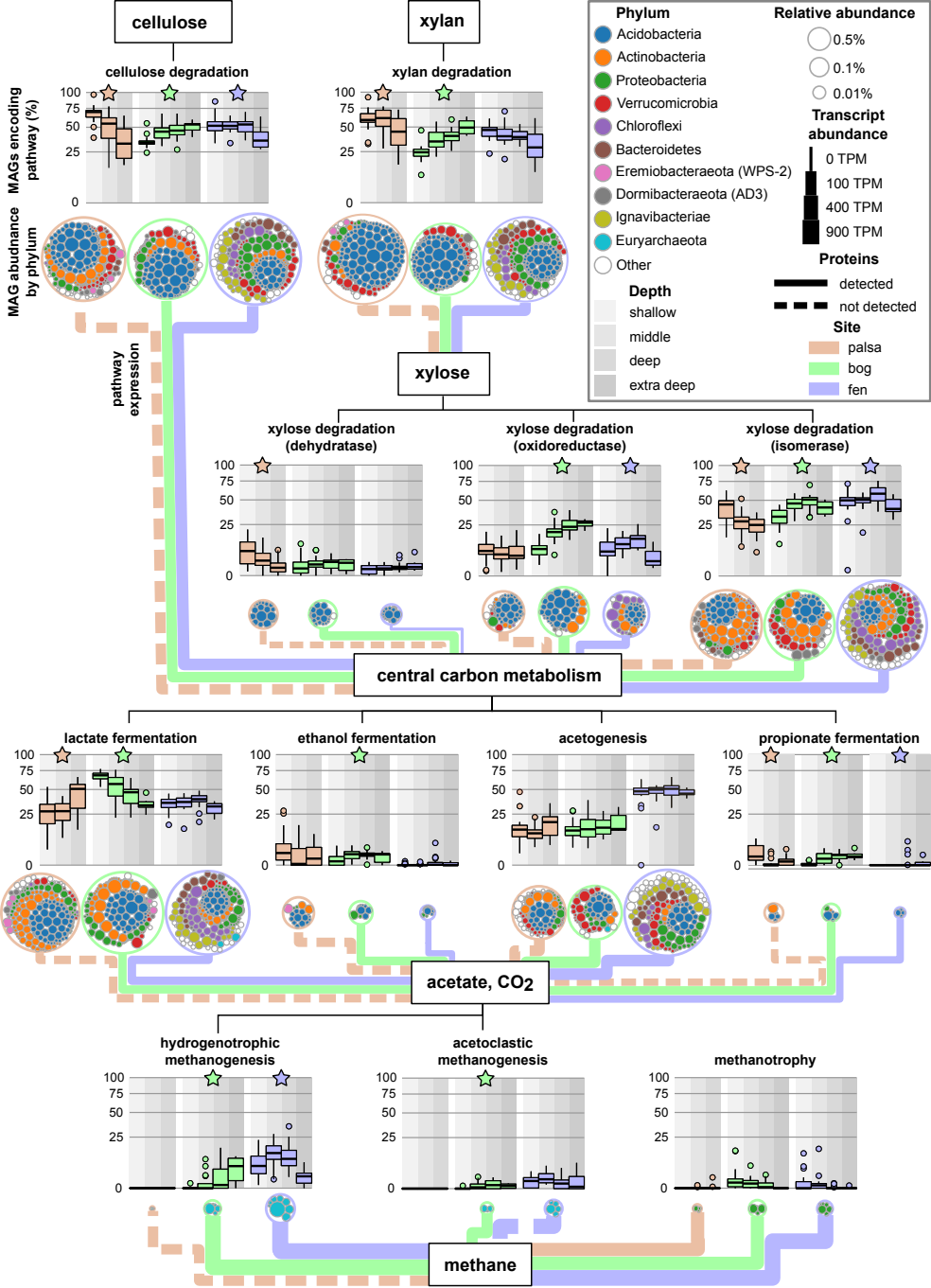
44 Payette, S., Delwaide, A., Caccianiga, M. & Beauchemin, M. Accelerated thawing of subarctic peatland permafrost over the last 50 years. *Geophysical Research Letters* **31** (2004).

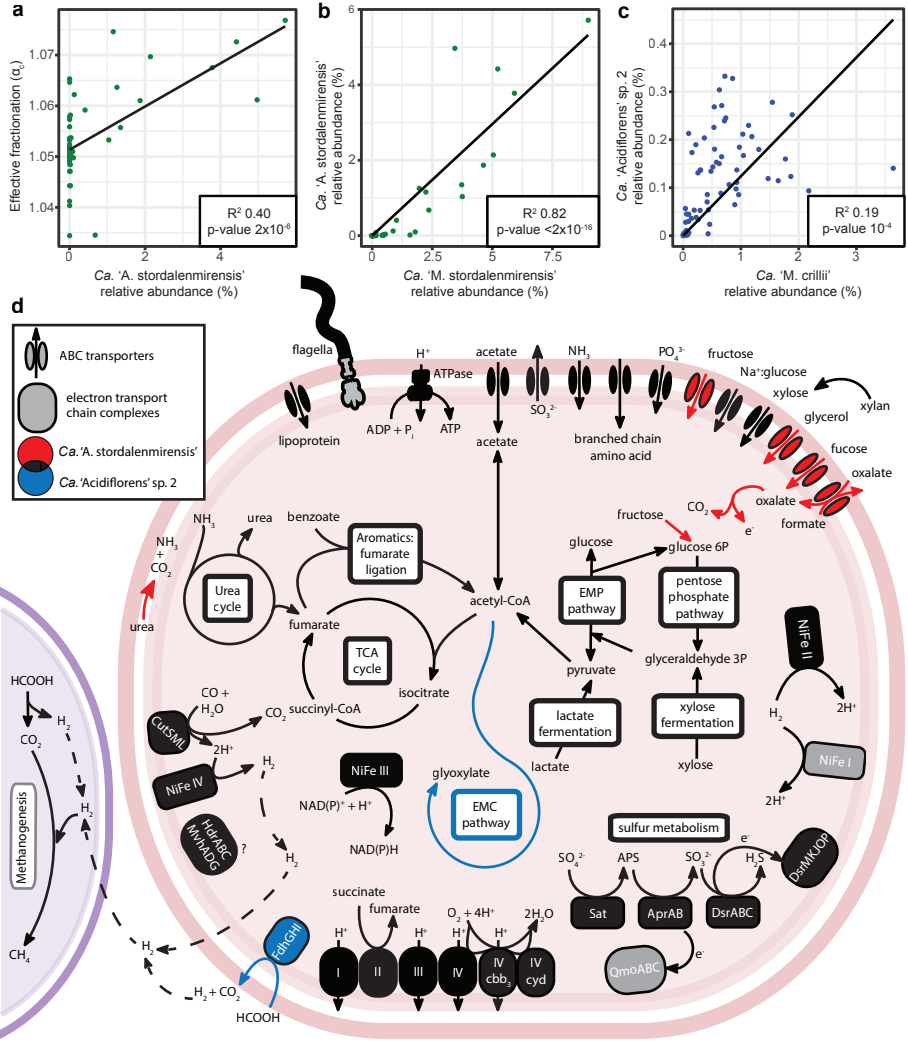
45 O'Donnell, J. A. *et al.* The effects of permafrost thaw on soil hydrologic, thermal, and carbon dynamics in an Alaskan peatland. *Ecosystems* **15**, 213-229 (2012).

46 Zoltai, S. Cyclic development of permafrost in the peatlands of northwestern Alberta, Canada. *Arctic and Alpine Research*, 240-246 (1993).
47 Quinton, W., Hayashi, M. & Chasmer, L. Permafrost-thaw-induced land-cover change in the Canadian subarctic: implications for water resources. *Hydrological Processes* **25**, 152-158 (2011).
48 Koven, C. D. *et al.* Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences* **108**, 14769-14774 (2011).
49 Melton, J. *et al.* Present state of global wetland extent and wetland methane modelling: conclusions from a model intercomparison project (WETCHIMP). *Biogeosciences* **10**, 753-788 (2013).
50 Whiticar, M. J., Faber, E. & Schoell, M. Biogenic methane formation in marine and freshwater environments: CO2 reduction vs. acetate fermentation—isotope evidence. *Geochimica et Cosmochimica Acta* **50**, 693-709 (1986).
51 Angly, F. E. *et al.* CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2**, 11 (2014).
52 Pielou, E. C. Shannon's formula as a measure of specific diversity: its use and misuse. *The American Naturalist* **100**, 463-465 (1966).
53 Oksanen, J. *et al.* The vegan package. *Community ecology package* **10**, 631-637 (2007).
54 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
55 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009).
56 Tange, O. Gnu parallel-the command-line power tool. *The USENIX Magazine* **36**, 42-47 (2011).
57 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).
58 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 Research* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).
59 Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics* **11**, 538 (2010).
60 Evans, P. N. *et al.* Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**, 434-438 (2015).
61 Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* **35**, D61-D65 (2007).
62 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature microbiology* (2017).
63 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490 (2010).
64 McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **6**, 610-618 (2012).
65 Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic acids research* **32**, 1363-1371 (2004).
66 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **44**, W242-W245 (2016).
67 Team, R. C. R language definition. *Vienna, Austria: R foundation for statistical computing* (2000).
68 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, btu153 (2014).
69 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, gkr367 (2011).

70 Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* **40**, W445-W451 (2012).
71 Bairoch, A. The ENZYME database in 2000. *Nucleic acids research* **28**, 304-305 (2000).
72 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60 (2015).
73 Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-D495 (2014).
74 Morozova, D. & Wagner, D. Stress response of methanogenic archaea from Siberian permafrost compared with methanogens from nonpermafrost habitats. *FEMS microbiology ecology* **61**, 16-25 (2007).
75 Lay, J.-J., Miyahara, T. & Noike, T. Methane release rate and methanogenic bacterial populations in lake sediments. *Water research* **30**, 901-908 (1996).
76 Noike, T. Upgrading of anaerobic digestion of waste activated sludge by thermal pretreatment. *Water Science and Technology* **26**, 857-866 (1992).
77 Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiological reviews* **60**, 439-471 (1996).
78 Costello, A. M., Auman, A. J., Macalady, J. L., Scow, K. M. & Lidstrom, M. E. Estimation of methanotroph abundance in a freshwater lake sediment. *Environmental microbiology* **4**, 443-450 (2002).
79 Baani, M. & Liesack, W. Two isozymes of particulate methane monooxygenase with different methane oxidation kinetics are found in Methylocystis sp. strain SC2. *Proceedings of the National Academy of Sciences* **105**, 10203-10208 (2008).
80 Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)* **28**, 3211-3217 (2012).
81 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841-842 (2010).
82 Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences* **131**, 281-285 (2012).
83 Guo, X. & Kristal, B. S. The use of underloaded C 18 solid-phase extraction plates increases reproducibility of analysis of tryptic peptides from unfractionated human plasma. *Analytical biochemistry* **426**, 86-90 (2012).
84 Kim, S. & Pevzner, P. A. Universal database search tool for proteomics. *Nature communications* **5**, 5277 (2014).
85 Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology* **30**, 918-920 (2012).
86 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).
87 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).
88 Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Research*, gky174-gky174, doi:10.1093/nar/gky174 (2018).
89 Racine, J. S. RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics* **27**, 167-172 (2012).
90 Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**, D447-456, doi:10.1093/nar/gkv1145 (2016).
91 Bertin, P. N. *et al.* Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta-and proteo-genomics. *The ISME journal* **5**, 1735-1747 (2011).


relative abundance (%)
0.005
0.01
0.05
0.1
0.5
1
5
10
100
a
Palsa
Bog
Fen
Shallow
Middle
Deep
Shallow
Middle
Deep
E
Shallow
Middle
Deep
E
b
Bacteria
Acidobacteria (364, 30%)
Actinobacteria
Dormibacteraeota (AD3) (47, 56%)
Armatimonadetes
Caldiserica
Bacteroidetes
Chlamydiae
Cyanobacteria
Chloroflexi
Elusimicrobia
FCPU426 (6, 70%)
Fibrobacteres
Firmicutes
Nitrospirae
Patescibacteria
Omnitrophica
Planctomycetes
Gemmatimonadetes (24, 26%)
Ignavibacteriae (50, 21%)
Proteobacteria
Spirochaetes
TM6
Verrucomicrobia
WOR-3
Eremiobacteraeota (WPS-2) (53, 57%)
Archaea
Bathyarchaeota
Euryarchaeota
Thaumarchaeota
c
Shannon diversity
4.6
4.4
4.2
4.0
3.8
3.6
3.4


cellulose
xylan
cellulose degradation
xylan degradation
MAGs encoding pathway (%)
MAG abundance by phylum
pathway expression
Phylum
Acidobacteria
Actinobacteria
Proteobacteria
Verrucomicrobia
Chloroflexi
Bacteroidetes
Eremiobacteraeota (WPS-2)
Dormibacteraeota (AD3)
Ignavibacteriae
Euryarchaeota
Other
Relative abundance
0.5%
0.1%
0.01%
Transcript abundance
0 TPM
100 TPM
400 TPM
900 TPM
Proteins
detected
not detected
Depth
shallow
middle
deep
extra deep
Site
palsa
bog
fen
xylose
xylose degradation (dehydratase)
xylose degradation (oxidoreductase)
xylose degradation (isomerase)
central carbon metabolism
lactate fermentation
ethanol fermentation
acetogenesis
propionate fermentation
acetate, CO2
hydrogenotrophic methanogenesis
acetoclastic methanogenesis
methanotrophy
methane


a
Effective fractionation ($\alpha_c$)
1.07
1.06
1.05
1.04
0
2
4
Ca. 'A. stordalenmirensis' relative abundance (%)
$R^2$ 0.40
p-value $2x10^{-6}$
b
Ca. 'A. stordalenmirensis' relative abundance (%)
6
4
2
0
0
2.5
5
7.5
Ca. 'M. stordalenmirensis' relative abundance (%)
$R^2$ 0.82
p-value $<2x10^{-16}$
c
Ca. 'Acidiflorens' sp. 2 relative abundance (%)
0.4
0.3
0.2
0.1
0
0
1
2
3
Ca. 'M. crillii' relative abundance (%)
$R^2$ 0.19
p-value $10^{-4}$
d
ABC transporters
electron transport chain complexes
Ca. 'A. stordalenmirensis'
Ca. 'Acidiflorens' sp. 2
flagella
lipoprotein
$H^+$
ATPase
$ADP + P_i$
ATP
acetate
acetate
$SO_3^{2-}$
$NH_3$
branched chain amino acid
$PO_4^{3-}$
fructose
$Na^+$:glucose
xylose
xylan
glycerol
fucose
oxalate
oxalate
formate
$CO_2$
$e^-$
fructose
glucose
glucose 6P
EMP pathway
pentose phosphate pathway
glyceraldehyde 3P
pyruvate
lactate fermentation
lactate
xylose fermentation
xylose
NiFe II
$H_2$
$2H^+$
NiFe I
$2H^+$
$NH_3$
urea
benzoate
Aromatics: fumarate ligation
$NH_3 + CO_2$
urea
Urea cycle
fumarate
acetyl-CoA
TCA cycle
isocitrate
succinyl-CoA
$CO + H_2O$
$CO_2$
CutSMB
$2H^+$
$H_2$
NiFe IV
HdrABC MvhADG
?
$H_2$
NiFe III
$NAD(P)^+ + H^+$
NAD(P)H
glyoxylate
EMC pathway
sulfur metabolism
$SO_4^{2-}$
APS
$SO_3^{2-}$
$H_2S$
$e^-$
DsrMKJOP
Sat
AprAB
DsrABC
QmoABC
$e^-$
succinate
fumarate
$H^+$
$H^+$
$O_2 + 4H^+$
$2H_2O$
$H^+$
I
II
III
IV
IV cbb3
IV cyd
FdhGH
HCOOH
$H_2 + CO_2$
$H_2$
HCOOH
$CO_2$
$H_2$
Methanogenesis
$H_2$
$CH_4$