

Microbial polyphenol metabolism is part of the thawing permafrost carbon cycle

Received: 8 December 2023

Accepted: 2 April 2024

Published online: 28 May 2024

 Check for updates

Bridget B. McGivern¹, Dylan R. Cronin^{2,3}, Jared B. Ellenbogen¹, Mikayla A. Borton¹, Eleanor L. Knutson⁴, Viviana Freire-Zapata⁵, John A. Bouranis⁵, Lukas Bernhardt⁶, Alma I. Hernandez⁶, Rory M. Flynn¹, Reed Woyda¹, Alexandra B. Cory⁷, Rachel M. Wilson⁸, Jeffrey P. Chanton⁸, Ben J. Woodcroft⁹, Jessica G. Ernakovich⁶, Malak M. Tfaily⁵, Matthew B. Sullivan^{2,3}, Gene W. Tyson⁹, Virginia I. Rich², Ann E. Hagerman⁴ & Kelly C. Wrighton¹✉

With rising global temperatures, permafrost carbon stores are vulnerable to microbial degradation. The enzyme latch theory states that polyphenols should accumulate in saturated peatlands due to diminished phenol oxidase activity, inhibiting resident microbes and promoting carbon stabilization. Pairing microbiome and geochemical measurements along a permafrost thaw-induced saturation gradient in Stordalen Mire, a model Arctic peatland, we confirmed a negative relationship between phenol oxidase expression and saturation but failed to support other trends predicted by the enzyme latch. To inventory alternative polyphenol removal strategies, we built CAMPER, a gene annotation tool leveraging polyphenol enzyme knowledge gleaned across microbial ecosystems. Applying CAMPER to genome-resolved metatranscriptomes, we identified genes for diverse polyphenol-active enzymes expressed by various microbial lineages under a range of redox conditions. This shifts the paradigm that polyphenols stabilize carbon in saturated soils and highlights the need to consider both oxic and anoxic polyphenol metabolisms to understand carbon cycling in changing ecosystems.

Permafrost stores an estimated 50% of global soil carbon^{1,2}, nearly twice the amount of carbon in the atmosphere³. Increasing global temperatures are thawing permafrost and risk destabilizing this carbon⁴. Newly thawed carbon may be decomposed by resident microbial communities, yielding carbon dioxide (CO₂) or methane (CH₄), greenhouse gases that can accelerate climate warming^{5,6}. Therefore, understanding microbial carbon processing in these climatically vulnerable habitats is critical to accurately project global warming feedbacks.

For decades, soil carbon stability in these systems was thought to be governed by the enzyme latch theories^{7–9} which propose that polyphenols inhibit microbial decomposition under anoxic conditions to ‘dead end’ soil carbon cycling. Polyphenols are a chemically diverse and abundant¹⁰ group of plant-derived compounds, spanning more than 10,000 chemical formulae and several structural families (Supplementary Fig. 1), including polymers (for example, tannins), monomers (for example, flavonoids) and simple phenols (for example, phenolic

¹Department of Soil and Crop Science, Colorado State University, Fort Collins, CO, USA. ²Department of Microbiology, The Ohio State University, Columbus, OH, USA. ³Center of Microbiome Science, The Ohio State University, Columbus, OH, USA. ⁴Department of Chemistry and Biochemistry, Miami University, Oxford, OH, USA. ⁵Department of Environmental Science; University of Arizona, Tucson, AZ, USA. ⁶Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA. ⁷Department of Environmental Sciences, Emory University, Atlanta, GA, USA. ⁸Department of Earth Ocean and Atmospheric Sciences, Florida State University, Tallahassee, FL, USA. ⁹Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Woolloongabba, Queensland, Australia.

✉e-mail: Kelly.wrighton@colostate.edu

acids)¹¹. Despite this vast chemical diversity, the enzyme latch theory recognizes oxygen-requiring phenol oxidases (POs) as being solely responsible for eliminating polyphenolic and phenolic compounds^{7,9} (Supplementary Note 1). Under this theory, saturated soil conditions (such as those post permafrost thaw) would render POs inactive, resulting in accumulated polyphenols that restrict microbial carbon decomposition, ultimately halting microbial soil carbon decomposition to stabilize soil carbon. Several studies have found conflicting support for parts of the enzyme latch theory^{12–15} or provided alternative biogeochemical explanations for results in the cascade¹⁶. However, even in these refuting studies, the primacy of POs in polyphenol degradation was taken for granted, leaving the possibility of alternative microbial metabolic strategies unexplored.

The microbial assumptions underlying the enzyme latch theory are not aligned with concepts from modern microbiome science. First, plants have been producing polyphenols for hundreds of millions of years^{17,18}. Therefore, it is likely that diverse mechanisms for utilization, resistance and tolerance have evolved in microbial communities often exposed to polyphenols¹⁹. Second, the assumption that a single enzyme type transforms a broad class of compounds, and under strict environmental conditions, contrasts examples of enzyme specificity at environmental²⁰, compound²¹ and even stereochemical levels²². Indeed, biochemical investigations across ecosystems such as the human gut have uncovered multiple non-PO microbial enzymes that degrade diverse polyphenols, including examples under anoxia²³. Critically, the use of these enzymes by peatland and even terrestrial microbiomes has yet to be demonstrated *in situ*.

Here we sought to revisit the fundamental assumptions of the enzyme latch theory and to more broadly describe microbial polyphenol transformations expressed in peatland soils. We selected a model Arctic permafrost peatland, Stordalen Mire^{5,24}, where the enzyme latch theory has been suggested to mediate carbon storage^{25–27}. At this site, natural thaw created three distinct habitats from dry, intact permafrost *palsa* through a partially thawed bog with a fluctuating water table, to a fully thawed and saturated fen (Supplementary Fig. 2). To track microbial polyphenol metabolism across this thaw gradient, we obtained paired genome-resolved metatranscriptome, metabolite and geochemical data from peat cores taken from the *palsa*, bog and fen in July 2016. Leveraging this dataset, we provide a revised view of polyphenols in microbially catalysed carbon cycling in these climate-threatened habitats.

Results

Challenges to the enzyme latch in Stordalen Mire

First, we wanted to establish whether relationships put forth in the enzyme latch theories held true in Stordalen Mire. If an enzyme latch was controlling microbial carbon cycling, we expected to observe: (1) a negative relationship between PO expression and water saturation, (2) a negative relationship between PO expression and polyphenol concentrations, (3) a negative relationship between polyphenols and extracellular hydrolase enzymes and (4) a negative relationship between polyphenols and porewater CO₂ and CH₄ (Fig. 1a). We assessed these variables with the traditional methods used to support the enzyme latch theory^{7,12}. To ensure that interpretation was not hampered by methodological constraints, we paired the traditional methods with high-resolution, multi-omic microbiome methods.

To test the first prediction, we analysed differences in assayed PO activity and metatranscriptome expression of PO genes between unsaturated and saturated peat samples. Using a traditional phenol oxidase assay, we saw no significant difference in PO activity between unsaturated and saturated samples (Fig. 1b(i) and Supplementary Data 1). However, from the metatranscriptome data, we detected significantly less PO gene expression in saturated samples than in unsaturated samples, mirroring the expected relationship of these oxygen-requiring genes (Fig. 1c(i) and Supplementary Data 1). These conflicting results

between traditional enzyme assay and our field gene expression data could be due to the convention of conducting the PO assay under oxic conditions in the laboratory instead of maintaining the field relevant redox conditions¹², suggesting that metatranscriptome expression is a more sensitive measure of PO expression under field conditions.

For the second prediction, in contrast to theoretical expectations, we did not observe a significant negative relationship between POs and polyphenol concentrations using either traditional or multi-omic methods (Fig. 1(bii),c(ii) and Supplementary Data 1). Similarly, there were no correlations between polyphenols and assayed enzyme activity or glycoside hydrolase metatranscriptome gene expression (Fig. 1b(iii),c(iii), and Supplementary Fig. 3 and Data 1). Lastly and notably, in contrast to the primary tenet of the enzyme latch theory, we found that polyphenol abundance was positively correlated with porewater CO₂ (although not CH₄) concentrations, indicating that polyphenol metabolism could contribute to peatland soil respiration (Fig. 1b(iv),c(iv) and Supplementary Data 1).

Collectively, traditional and multi-omics methods failed to support relationships expected by the enzyme latch theory at key steps in the biogeochemical cascade. This was consistent whether Stordalen Mire's peatlands were assessed collectively (Fig. 1) or by habitat (that is, *palsa*, bog and fen; Supplementary Fig. 4 and Data 1), suggesting that the enzyme latch theory is not a suitable descriptor of this peatland's carbon processing dynamics at any scale. Further, the lack of relationship between polyphenol oxidases and polyphenol content suggests that POs alone do not control the fate of these compounds along this thaw gradient. Given this and the positive relationship observed between polyphenol abundance and porewater CO₂ concentrations, our data indicate that polyphenols do not inhibit microbial carbon metabolism and may even contribute to respiration *in situ*.

Polyphenol transformations expressed across Stordalen Mire

We next sought to investigate the avenues of microbial polyphenol metabolism active in Stordalen Mire. Although the enzyme latch theory has long suggested that POs were the sole microbial enzymes controlling polyphenol degradation in soil systems⁷, biochemical investigations across microbiomes from diverse ecosystems have biochemically characterized microbial enzymes that act upon distinct polyphenol types, under a range of redox conditions and for diverse metabolic ends^{23,28}. Unfortunately, these annotations were not translated into annotation frameworks, thus impeding their broader extension to other microbiomes.

Leveraging these studies, we developed an open-source annotation tool called CAMPER (Curated Annotations for Microbial Polyphenol Enzymes and Reactions) that annotates and summarizes polyphenol transformation potential from assembled gene data²⁹. This tool searches 41 custom (Supplementary Table 1) and 234 database-derived homology-based gene family annotations against user input sequence files. CAMPER assigns genes to a substrate ontology on the basis of established polyphenol chemical classes³⁰ as well as aggregates this content into 100 transformation pathways (Supplementary Data 2). In addition, CAMPER pathways were coded for their oxygen dependence on the basis of previous biochemical knowledge. CAMPER pathways were also placed into a trophic degradation hierarchy on the basis of their transformation of polymeric polyphenols, monomeric polyphenols, or simple phenols. Finally, any additional information needed to manually curate the CAMPER annotations is provided, including gene cluster organization and expected secretion status. CAMPER compiles existing biochemical knowledge into a chemical framework to make polyphenol metabolisms more accessible to microbiome scientists.

To identify potential polyphenol transformation strategies used by the Stordalen Mire microbiome, we used CAMPER to annotate a database of 1,864 metagenome-assembled genomes (MAGs) recovered from field-derived metagenomes taken over 8 years across the

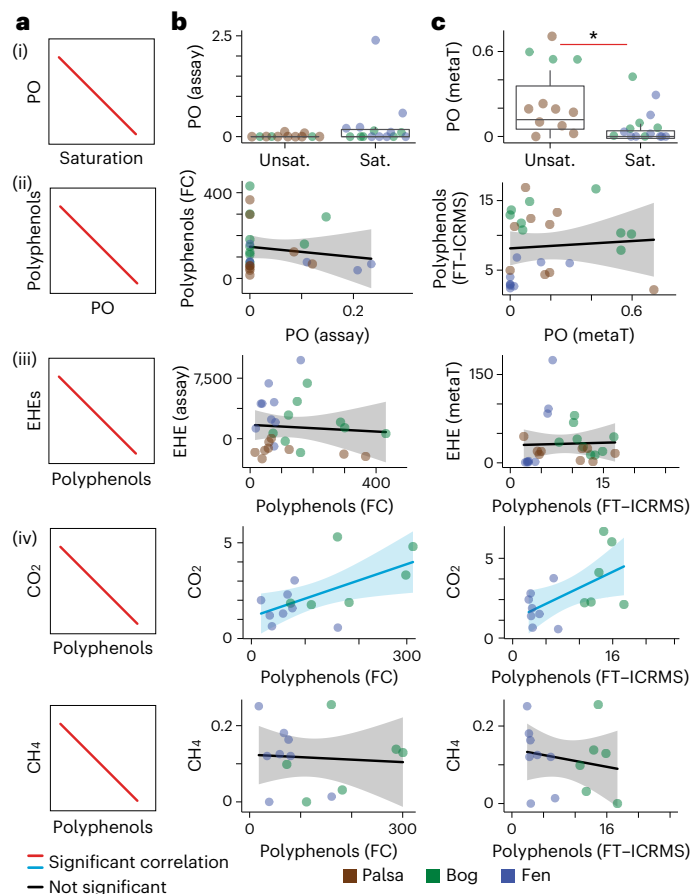


Fig. 1 | Multimethod investigation of the enzyme latch theory. **a**, The expected relationships proposed by the enzyme latch theory, with relationships labelled (i)–(iv) as in the text. **b, c**, The matched observed relationships using ‘Traditional’ assay methods (**b**) versus ‘omics and high-resolution methods’ (**c**). In **b(i)** and **c(i)**, the lower and upper boxplot edges represent the 25th and 75th percentiles, respectively, and the middle line is the median. The whiskers extend from the median to $1.5 \times$ the interquartile range. $*P = 0.005848$, Wilcoxon rank-sum test. In **b(ii)**–(iv) and **c(ii)**–(iv), the linear trendlines for significant Pearson correlations (Benjamini–Hochberg $P_{\text{adj}} < 0.05$) are coloured blue (positive correlation) and red (negative correlation). The shaded regions correspond to the 95% confidence intervals. Water saturation is used as a proxy for oxygen availability, with unsaturated (Unsat.) corresponding to high oxygen, and saturation (Sat.) corresponding to low/no oxygen. Phenol oxidase activity is given by phenol oxidase assay (PO assay, nmol activity $\text{g}^{-1} \text{h}^{-1}$) and the summed metatranscriptome expression of phenol oxidase genes from MAGs in the metatranscriptome (see Methods, PO metaT, geTMM). Polyphenol content is given by Folin–Ciocalteu phenolics (FC polyphenols, mg methyl-gallate equivalents per dry g soil) and polyphenol-like compounds identified in FT-ICRMS (% polyphenols; Supplementary Fig. 11). Extracellular hydrolase enzymes (EHE) are given by beta-glucosidase activity (EHE assay, nmol activity $\text{g}^{-1} \text{h}^{-1}$) and the summed expression of genes encoding glycoside hydrolases from MAGs in the metatranscriptomes (EHE metaT, geTMM). This is representative of relationships observed for other enzyme types (see Supplementary Fig. 3). Porewater CO_2 and CH_4 are given in concentrations (mM). Individual correlation coefficients, P values and sample sizes are given in Supplementary Data 1.

site (Supplementary Data 2). We used this genomic framework to contextualize a metatranscriptome and metabolome dataset derived from triplicate palsa, bog and fen cores taken from July 2016, with cores divided at three depths (Supplementary Data 3 and 4). As preliminary evidence for polyphenol transformations in these soils, metabolomics (liquid chromatography with tandem mass spectrometry (LC–MS/MS)) identified 35 polyphenols across six families. In further support of active polyphenol metabolisms in situ, the metatranscriptome data

uncovered 58 expressed polyphenol transformation pathways from across seven polyphenol families (Fig. 2, and Supplementary Figs. 5 and 6). Four of the detected polyphenol metabolites were substrates or products in expressed CAMPER pathways (Supplementary Fig. 5). We observed habitat- and depth-specific metabolite and gene expression patterns, probably due to differences in plant and microbial community species between habitats^{5,31}. Importantly, unlike PO alone (Fig. 1b(ii), c(ii)), the collective expression of these 58 metatranscriptome pathways was significantly and positively related to polyphenol content (Supplementary Fig. 7). These findings illustrate the microbial metabolic plasticity overlooked by the enzyme latch theory which focuses on a single enzymatic type.

Due to varying water saturation conditions at the time of sampling (with all palsa depths and surface bog being dry, and deeper bog and fen being water saturated), we expected to observe an associated variation in oxic and anoxic polyphenol-transformation gene expression (Fig. 2a and Supplementary Fig. 2). In accordance with water table depth across these sites, the relative expression of genes encoding enzymes for anoxic polyphenol transformations was significantly higher in the saturated samples (Supplementary Fig. 7). This contrasts with the enzyme latch theory, as waterlogging does not appear to turn off microbial polyphenol metabolism.

Next, we inventoried polyphenol metabolism across the degradation hierarchy (polymeric polyphenol, monomeric polyphenol and phenolic/benzoic transformations). For polymers, genes encoding PO (AA1) were most enriched in bog surface-layer metatranscriptomes, consistent with the enzyme’s oxygen dependence (Fig. 2a no. 22 and Fig. 2b). In addition, we detected expression of genes encoding three non-PO enzymes reported to depolymerize polymeric polyphenols: peroxidase (AA2, Fig. 2a no. 39 and Fig. 2b), vanillyl-alcohol oxidase (AA4, Fig. 2a no. 35 and Fig. 2b) and benzoquinone reductase (AA6, Fig. 2a no. 47 and Fig. 2b)^{32–34}. While AA2 and AA4 expression was enriched in the unsaturated bog samples, AA6 expression was enriched in the saturated fen samples and increased with depth (Fig. 2a). As bacterial AA6 enzymes have been suggested to function in anoxic polyphenol depolymerization³⁴, their expression in the fen further refutes the enzyme latch assumption that polyphenol degradation is halted under water-saturated conditions. Rather, our data illustrate that peatland microbiomes express genes for multiple enzymes beyond phenol oxidases that operate across a range of redox conditions.

For monomeric polyphenols and phenolic/benzoic acids, we observed a trend similar to that of polymers where at a polyphenol family level, distinct transformation pathways were expressed depending on the redox environments. For example, the oxic pathways for flavonoids were enriched in the palsa and bog (Fig. 2a nos. 6 and 21 and Fig. 2c), and the anoxic flavonoid pathway was enriched in the water-saturated fen (Fig. 2a no. 55 and Fig. 2c). Furthermore, the oxic degradation pathway for benzoic acid was enriched in the palsa (Fig. 2a no. 3 and Fig. 2d), while three anoxic degradation pathways were enriched in the fen (Fig. 2a nos. 45, 49 and 50 and Fig. 2d). In all, this is important because these distinct transformation pathways associated with varying water saturation across the site lead to different metabolic outcomes. Taken together, we put forward the idea that redox state is an important environmental filter, not an ‘on/off’ switch, that alters the fate and metabolic consequences of polyphenols.

Our data further illuminated the myriad ways that polyphenols can contribute to microbial metabolism. Beyond supporting microbial metabolism as an electron donor or carbon source, the bog-enriched caffeic acid reductase system instead uses the polyphenol caffeic acid to accept electrons (Fig. 2a no. 37 and Fig. 2e)³⁵. In this pathway, caffeic acid is activated to caffeoyl-CoA and subsequently reduced to hydrocaffeoyl-CoA, which transfers the CoA to another caffeic acid in an energy-saving loop. The reduction step generates NAD^+ and reduces ferredoxin, which fuel the Rnf complex and a hydrogenase²⁸. Interestingly, caffeic acid metabolite abundance is positively related to caffeic

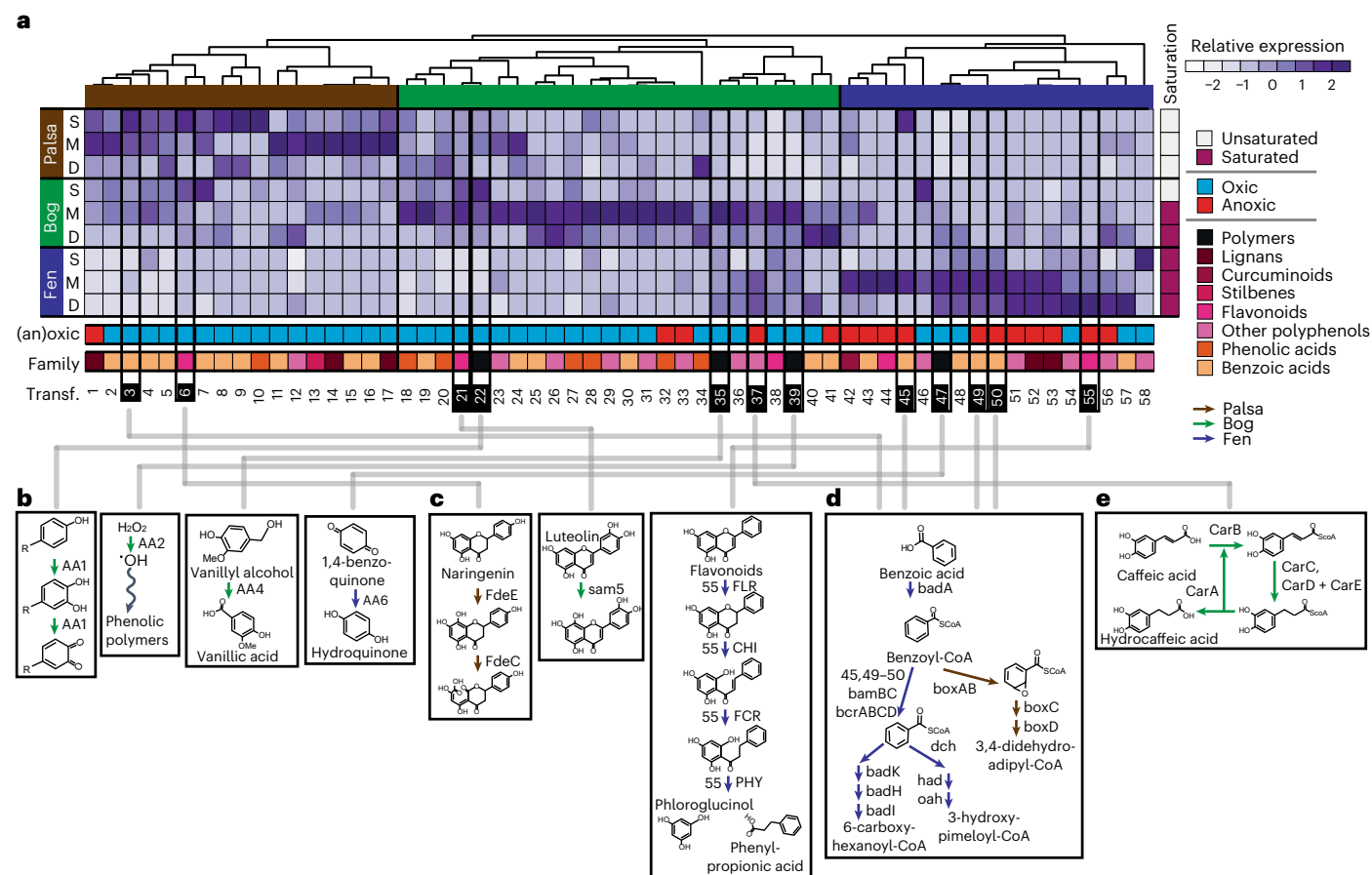


Fig. 2 | Polyphenol transformations expressed in Stordalen Mire.

a, Hierarchical clustering of the relative metatranscriptome expression of polyphenol transformations across habitats and depths (rows: S, surface; M, middle; D, deep depths). Here, relative metatranscriptome expression is displayed as the z-score of the average of the summed expression ($n = 3$, geTMM) of genes in each pathway (columns). The saturation column to the right of the heat map denotes samples as saturated (red) if they were at or below the water table depth, or unsaturated (white) if they were above the water table (Supplementary Data 1 and Fig. 2). Hierarchical clustering of pathways revealed palsa (brown), bog (green) and fen (blue) specific transformations, indicated by top-row colour. The oxygen requirements for transformations are shown in

the second bottom-most row for oxic (blue) and anoxic (red) transformations. The family of polyphenols that each pathway acts upon is coloured at the bottom-most row: black (polymers), pink shades (monomers) and orange shades (phenolic/benzoic acids). Column numbers correspond to transformations; see Supplementary Data 2 for more detail on each transformation. Transformations mentioned in the text are enclosed in a black box, and the numbers are highlighted black in the heat map to match the text reference, and the reactions are shown in **b–e**. Arrow colours correspond to the habitat cluster for the transformation: palsa (brown), bog (green) and fen (blue). In no. 39, the wavy arrow indicates that the hydroxyl radical diffuses away to act on phenolic polymers. All pathways are shown in Supplementary Fig. 6.

acid reductase gene expression in the fen (Supplementary Fig. 5), providing additional evidence that this metabolism may be occurring. This metabolism expands the role of polyphenols in soil microbiomes to alternative energy-generating mechanisms.

Collectively, our metatranscriptome data paint a complex view of microbiome–polyphenol interactions in peatlands, supporting the importance of polyphenols to peatland carbon cycling despite being historically overlooked as substrates. While the enzyme latch theory posited that strict redox conditions limit polyphenol transformations in peat, we revealed the potential for redox-tunable transformation strategies that enable microorganisms to use polyphenols for diverse metabolic ends.

Diverse and talented polyphenol-active lineages

Next, we wanted to catalogue the microorganisms encoding and expressing these polyphenol transformations. Illustrating the prevalence of these transformations in the thawed permafrost microbiome, only one of the 43 phyla captured by the 1,864 MAGs lacked appreciable polyphenol or subsequent phenolic transformations (the 58 Patescibacteria MAGs, Fig. 3a). MAGs from the Acidobacteriota, Proteobacteria,

Actinobacteriota and Eremiobacteriota were particularly enriched in polyphenol metabolism, encoding on average 10 transformations per genome (Fig. 3a and Supplementary Fig. 8). Further, expression of these transformations spanned MAGs across the phylogenetic tree, in all three habitats (Fig. 3a and Supplementary Fig. 9).

Considering the broad phylogenetic range of the MAGs encoding and expressing these transformations, we next identified the important polyphenol-active lineages in Stordalen Mire. To do this, we defined two metrics of importance derived from the metatranscriptome data: (1) polyphenol talent, the number of different pathways expressed and (2) polyphenol dominance, the relative contribution to active polyphenol metabolism in Stordalen Mire (see Methods and Supplementary Fig. 10). We assessed these metrics for MAGs aggregated at the genus level ($n = 302$ genera) and identified 50 microbial genera spanning 11 phyla that exhibited polyphenol talent or dominance (Fig. 3b and Supplementary Fig. 11). While some of these were well characterized polyphenol and aromatic-degrading genera such as *Novosphingobium*³⁶, 71% of these genera were undescribed lineages (that is, alphanumeric names), highlighting the power of genome-resolved analyses to provide metabolic contexts for uncultured or poorly characterized lineages.

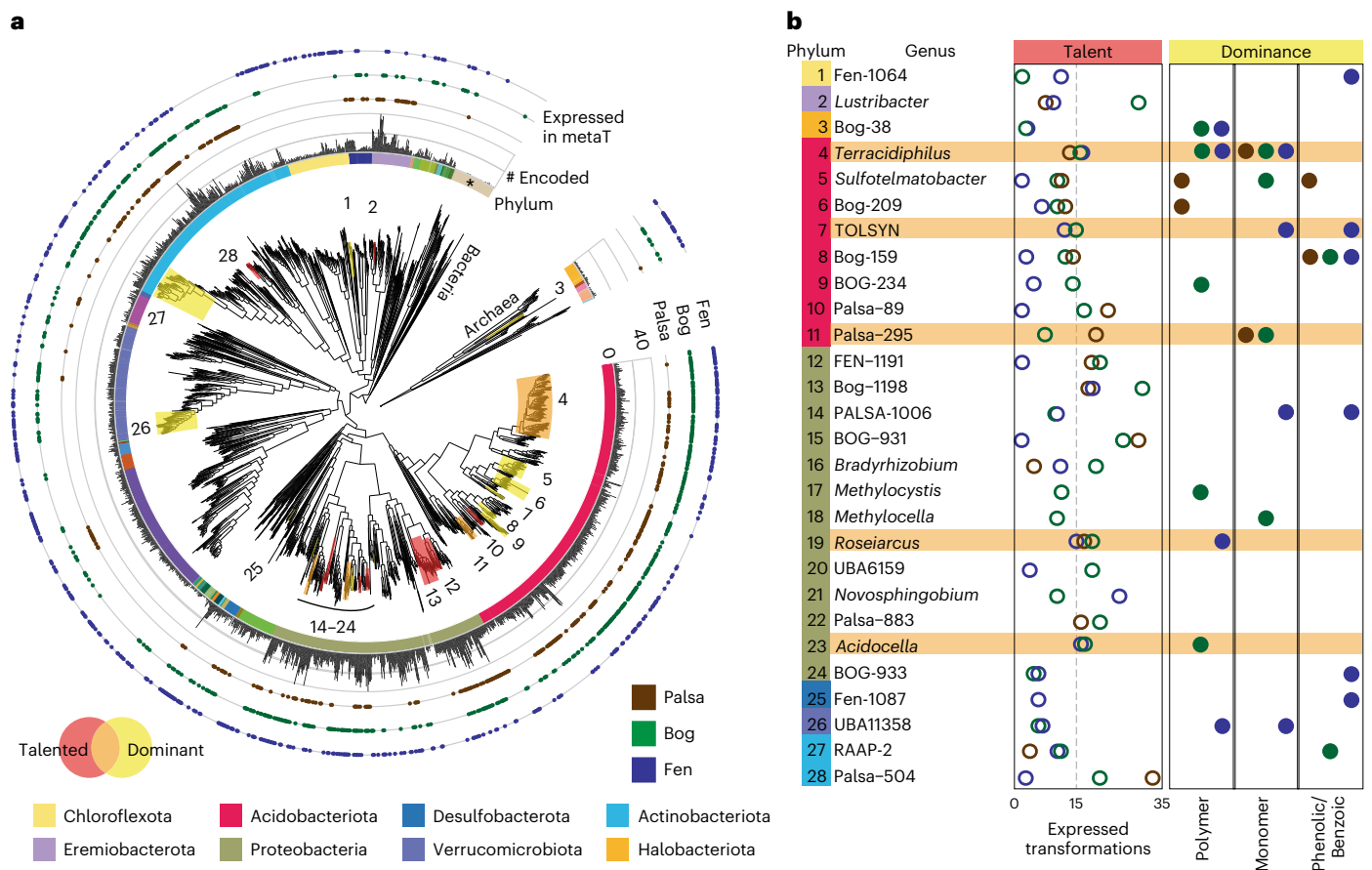


Fig. 3 | Polyphenol transformation potential encoded and expressed across Stordalen Mire MAGs. a, Phylogenetic trees of the 1,864 Stordalen Mire MAGs constructed using the GTDB 120 bacterial and 53 archaeal gene sets³³, with the Patescibacteria and Micrarchaeota as outgroups, respectively. The inner multicoloured ring corresponds to phylum, with the Patescibacteria (in beige) indicated by an asterisk (*). The middle ring displays a bar chart of the number of polyphenol transformations encoded by each MAG (range 0–40). The outer rings indicate whether a given MAG expressed at least one polyphenol transformation in the palsa, bog or fen metatranscriptomes by the presence of a dot. Clades

of ‘polyphenol talented’ genera are highlighted in red, polyphenol dominant genera are highlighted in yellow and genera that meet both definitions are highlighted in orange (bottom left legend). Numbering at clade tips corresponds to genus names in **b**. **b**, Plots showing the number of unique polyphenol transformation pathways expressed per genus in each habitat (left), and whether that genus contributed at least 10% of polymer, monomer or phenolic/benzoic acid transformations in any habitat (right). A dashed line at 15 expressed transformations delineates talented genera (Supplementary Fig. 10). Genera that are both talented and dominant are highlighted in orange.

An additional six genera were both talented and dominant: the Acidobacterial genera *Terracidiphilus* and Palsa-295, the Actinobacterial genus Palsa-504, and the Proteobacterial genera *Acidocella*, FEN-1191 and Bog-1198 (Fig. 3b). To the best of our knowledge, this represents presumably the first description of the polyphenol relevance of these lineages. Underscoring the relevance of these genera in the Stordalen Mire microbiome, MAGs from these six genera contributed between 6–65% of metatranscriptome expression across habitats and depth (Supplementary Fig. 9). Beyond Stordalen Mire, these genera have been reported to be important members of peatland microbiomes^{37,38}, for which we provide roles in carbon cycling.

Polyphenol-active lineages in the Stordalen carbon cycle

Finally, we wanted to reframe how polyphenols integrate into the Stordalen Mire carbon cycle. Our metatranscriptome data refined traditional conceptual models in permafrost^{5,25,39} by illustrating that polyphenols and phenolic acids may be actively used microbial substrates beyond polysaccharides and sugars (Supplementary Fig. 12). Therefore, to begin to place polyphenols in the carbon cycle, we analysed the metatranscriptome expression of MAGs in the six dominant and talented genera.

Illustrating the genus-level conservation of polyphenol talent in these genera, all MAGs expressed transformations at the polymeric,

monomeric and phenolic acid levels (Supplementary Fig. 12). MAGs belonging to the Palsa-295, Bog-1198 and Fen-1191 genera in particular expressed genes encoding enzymes for transformations of six different families of monomeric polyphenols (Supplementary Fig. 12 and Data 5). Furthermore, Palsa-295, Palsa-504, Fen-1191 and Bog-1198 members expressed genes for enzymes that decarboxylate benzoic acids, providing a specific example of how these compounds may directly contribute to CO₂ production (Fig. 1, and Supplementary Fig. 12 and Data 5). Consistent with expanding polyphenol metabolism across broader redox regimes, MAGs from all six genera had active respiratory lifestyles, expressing genes for enzymes in aerobic respiration, iron, nitrogen and/or sulfur reduction (Supplementary Fig. 12 and Data 5). Lastly, we note that in addition to polyphenol-transforming genes, MAGs from all six genera co-expressed genes for polysaccharide depolymerizing enzymes, and all but *Acidocella* MAGs expressed genes for enzymes in sugar degradation pathways. This further emphasizes the need to consider polyphenol transformations in addition to sugars and polysaccharides when assessing peatland carbon cycles. Collectively, these talented and dominant polyphenol-transforming genera illuminate how polyphenol transformations potentially contribute to Stordalen Mire biogeochemistry.

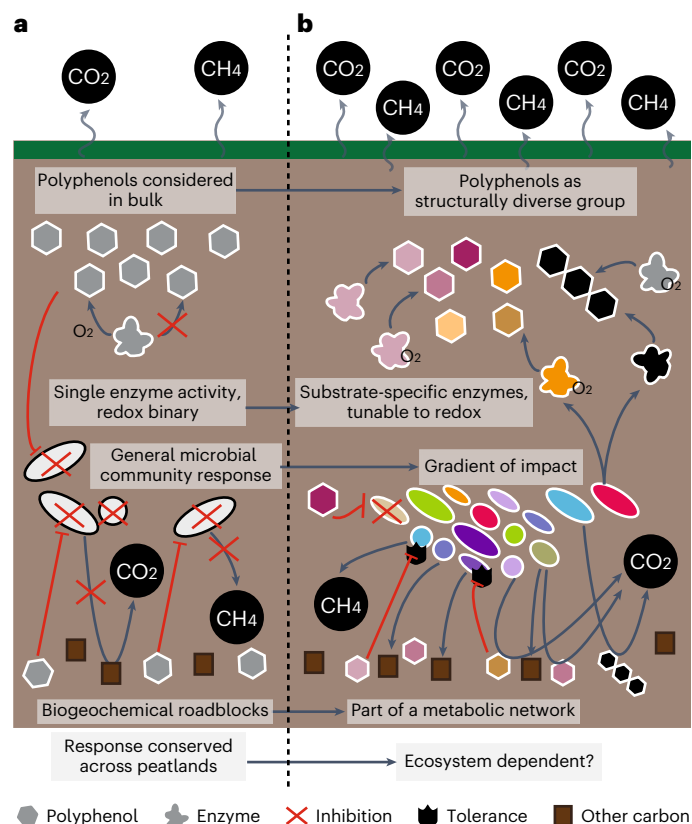


Fig. 4 | Conceptual reframing of polyphenols in peatland systems. a, The enzyme latch model posits that polyphenols can only be degraded by phenol oxidases under oxic conditions. Without oxygen, polyphenols accumulate and generally inhibit the microbial community, subsequently shutting down microbial metabolism, reducing CO₂ and CH₄ emissions. **b**, In this study, we propose a polyphenol-cognizant model that accounts for the fact that polyphenols are a diverse group of compounds with diverse metabolic impacts. Microbial communities encode and express multiple substrate-specific enzymes for diverse polyphenols, with strategies under both oxic and anoxic conditions. As a whole, the microbial community exhibits a gradient of responses to polyphenols, ranging from potential inhibition to stimulation. Due to this, polyphenols are integrated into carbon cycling networks, with susceptible organisms, tolerant organisms and stimulated organisms.

Beyond these six genera, several methane-cycling genera were identified as having capacity for polyphenol transformations, we posit likely for detoxification purposes. Notably, the methanogenic genus Bog-38 (*Candidatus Methanoflorens*), a critical methanogen at Stordalen Mire^{5,40}, was a dominant lineage in the bog and fen expressing genes for polymeric polyphenol-active enzymes (Fig. 3 and Supplementary Fig. 11), while *Methanosarcina* and the genus Fen-7 were dominant lineages expressing genes for polymeric polyphenol-active enzymes in the fen. This suggests that some methanogens may have resistance to polyphenols, contradicting the common assertion that methanogens are inhibited by polyphenols, either directly by cellular toxicity or indirectly through the polyphenols binding essential metals needed for enzyme biosynthesis⁴¹.

Of the five active Bog-38 MAGs, three expressed genes encoding benzoquinone reductase (AA6) and quercetin-2,3-dioxygenase (QueD, Fig. 2 no. 38 and Supplementary Fig. 12). Two of three *Methanosarcina* MAGs expressed genes for peroxidases and AA6, and one of two Fen-7 MAGs expressed genes for AA6. While AA6 has not been studied in methanogens, the enzyme has been suggested to catalyse transformations to detoxify phenolics³⁴. Furthermore, QueD uses oxygen to cleave the flavonoid quercetin, releasing carbon monoxide (CO)⁴².

This enzyme has also been suggested to serve as a polyphenol detoxification strategy⁴³. Intriguingly, the Bog-38 MAGs co-expressed genes for carbon monoxide dehydrogenase (CODH), potentially facilitating the use of the produced CO as an energy or carbon source^{44,45}. Beyond these polyphenol dominant genera, MAGs from two other methanogenic genera, *Methanobacterium_A* and *Methanobacterium_B*, also co-expressed genes for QueD and CODH (Supplementary Fig. 12). These findings lead us to hypothesize that polyphenol transformations by methanogens could contribute to resistance and may even directly support methanogenesis, although physiological experiments are needed to validate these suppositions.

Of additional relevance to methane cycling in these climate-critical habitats, two facultative methanotrophic genera were polyphenol dominant lineages in the bog: *Methylocella* and *Methylocystis* (Supplementary Figs. 11 and 12, and Data 5). MAGs from both genera exhibited nearly identical gene expression patterns, expressing genes for PO, QueD, the caffeic acid reduction system and 4-hydroxyphenylpyruvate dioxygenase (Hpp) for converting 4-hydroxyphenylpyruvate to homogentisic acid. *Methylocella* also expressed genes involved in homogentisic acid degradation, hinting at heterotrophic activity beyond methane oxidation, a supposition that has been supported for other carbon substrates^{46,47}. *Methylocystis* further expressed genes for AA6 and vanillyl-alcohol oxidase (AA4). Just as for the methanogens, while the physiological roles of these transformations are unclear and need to be experimentally investigated, the expression of these genes could enable polyphenol degradation, providing growth substrates or polyphenol detoxification. Furthermore, as these lineages are facultative methanotrophs, polyphenol transformation activity could impact CH₄ release from Stordalen Mire.

Discussion

Polyphenols have long been cited as controllers of peatland carbon cycles through the enzyme latch mechanism. The studies proposing and supporting this role have historically measured the impact of polyphenols on net outputs of carbon mineralization^{7–9,26}. We collected similar data, but also focused on the interplay between polyphenols and the microbiome in situ using multi-omics (Fig. 4). In all, our data refine the enzyme latch theory to be more reflective of modern microbiome science.

We first emphasize that polyphenols are not a bulk entity, but a diverse class of compounds with distinct impacts. We show that microbial communities engage with this chemical diversity, expressing many substrate-specific enzymes for polyphenols. This contrasts with the PO-centric enzyme latch mechanism, which acknowledges just one enzyme type. Furthermore, rather than a redox binary where anoxia inhibits polyphenol decomposition, we found that microbial communities tuned their gene expression strategies to habitat and redox conditions. Therefore, instead of a general antagonistic community response to polyphenols, we illustrated that the microbiome harbours members ranging from polyphenol-inhibited to stimulated. Importantly, we show that rather than shutting down metabolism for the whole microbial community, polyphenols are probably integrated into a larger carbon decomposition network where they serve as substrates, and their decomposition into smaller phenolics could directly fuel carbon dioxide and carbon monoxide production. Rather than relying on a single enzyme, we propose the talented microbiome model where soil microbiomes encode and express a cache of polyphenol metabolisms.

This model raises several questions that should be explored across terrestrial systems. First, the rates of polyphenol removal across the observed transformation pathways remain unknown. Second, the impact of diverse polyphenol degradation pathways for overall carbon storage, the key outcome of the enzyme latch, needs to be examined. Furthermore, the contributions of aerobic and anaerobic fungi to polyphenol degradation should be explored^{48,49}, as well as the direct

and indirect impacts of polyphenol metabolism on the methane cycle. We highlight CAMPER as a tool to facilitate detection of genes encoding polyphenol-active enzymes in genomic datasets, enabling researchers to begin addressing these questions in their datasets.

As permafrost carbon stores face uncertain fates with changing climates, we must not assume that portions of carbon are microbially unavailable without also considering the vast enzymatic diversity encoded in soil microbiomes. We highlight this work as a starting place for studying polyphenol dynamics in ecosystem carbon cycles in a post-enzyme latch world. Future assessments of peatland and permafrost microbial carbon cycles should consider the myriad ways that polyphenols support microbial metabolism to better understand and to predict ecosystem biogeochemistry.

Methods

Phenol oxidase

Here we refer to PO activities instead of a single PO enzyme to best reflect the biochemical understanding of this enzyme group (Supplementary Note 1). We measured PO activity through the L-3,4,-dihydroxy phenylalanine (L-DOPA) assay (see 'Enzyme assays'). To survey PO gene expression, we used the CAZyme family Auxiliary Activity 1 (AA1) to cover PO activity under Enzyme Commission number 1.10.3.-. Dioxygenases and monooxygenases were identified on their own.

Field site and sampling

Stordalen Mire is a thawing permafrost peatland in Northern Sweden (68° 22' N, 19° 03' E). The site has three main habitats along a permafrost thaw gradient, with distinct vegetation and thaw status: (1) palsa, overlaying permafrost, well-drained and dominated by woody plants and shrubs; (2) bog, intermediate thaw with variable water table depth characterized by a mixture of woody, herbaceous, sedge and moss plants, dominated by *Sphagnum* mosses and *Eriophorum vaginatum*; and (3) fen, fully thawed and waterlogged, dominated by *Eriophorum angustifolium* and *Carex* spp.

While MAGs were recovered from a longer period of field sampling (see 'MAG database construction'), all analyses described below were performed on samples collected in July 2016. Triplicate cores were taken from palsa, bog and fen sites (all within a 120 m radius and centred around an autochamber system) in Stordalen Mire using an 11-cm-diameter push corer. Cores were sectioned in the field at three depths: surface (1–5 cm), middle (10–14 cm) and deep (20–24 cm) (Supplementary Fig. 2). Each depth was split into separate samples for microbial and geochemical analyses in the field, immediately upon coring. For microbial analysis, peat from the interior of the core (that is, avoiding the outermost centimetre of core circumference) was added to Lifeguard buffer in the field (QIAGEN) at a 1:2.5 ratio, shaken vigorously, transported on ice to the research station and stored at –80 °C until shipment on dry ice, then returned to –80 °C storage until analysis. For geochemistry, the samples were stored in 50 ml falcon tubes at –20 °C.

Geochemical analyses

Before coring, porewater from each core site was collected using a perforated stainless-steel tubing inserted into the peat to the desired depth and extracted with an airtight syringe. No pore water was collected in the dry palsa or above the water table in the bog. Porewater was used to determine pH, and samples were acidified to measure porewater CO₂ and CH₄ using a flame-ionization-detector gas chromatograph equipped with a methanizer.

MAG database construction

To maximize the site-specific genomes available for this study, a database of 13,290 medium and high-quality MAGs was compiled from metagenomes taken from Stordalen Mire peat cores taken in field campaigns from 2010 to 2017 and from a stable isotope probing (SIP)

experiment performed on field peat added to labelled plant matter from locally co-occurring species. Notably, SIP metagenomes were used here purely to add site MAGs, not for analyses of that experiment. The methodology of MAG recovery is reported in ref. 50. MAG origin is given in Supplementary Data 3.

The 13,290 MAGs (with at least 70% completeness and less than 10% contamination, as determined through CheckM (v.1.0.12)⁵¹) were dereplicated using galah (v.0.3.0)⁵² with the following parameter set: –precluster-ani 90 –ani 97 –precluster-method finch. This resulted in 1,864 representative MAGs. MAGs were taxonomically annotated using the GTDB-tk (v.2.0)⁵³ classify workflow with the GTDB-Tk reference data v.r207. MAGs were annotated using DRAM (v.1.4.0)⁵⁴. MAG taxonomy and checkM statistics are available in Supplementary Data 3. DRAM annotations are available in Zenodo⁵⁵.

MAG metagenome abundance determination

To determine metagenome abundances of the 97% dereplicated MAG set, we first mapped trimmed metagenome reads to the MAG set using bowtie2 (v.2.4.5)⁵⁶ with the following settings: -D 10 -R 2 -N 1 -L 22 -i S,0,2,50. The output SAM file was converted to a sorted BAM using samtools (v.1.9)⁵⁷ and filtered using the reformat.sh script in the bbtools package using: idfilter=0.95 pairedonly=t primaryonly=t. MAG abundance was inferred from this using coverM (v.0.6.0)⁵⁸ as in ref. 32. MAG metagenome abundance is given in Supplementary Data 3.

RNA extraction

DNA and RNA were co-extracted using the Mobio PowerMax Soil DNA/RNA isolation kit (12966-10) with slight modifications. Briefly, sample vials were removed from the –80 °C freezer and thawed on ice. After thawing, 5–10 g of peat materials (preserved in Lifeguard soil preservation solution, QIAGEN) was added into the bead tubes in the kit, and the nucleic acids were extracted following the manufacturer protocol without the addition of beta-mercaptoethanol at the beginning. Reagents were proportionally increased to maintain the concentration and strength of solutions. An additional ethanol wash of nucleic acids-bound column was performed to further wash out impurities. The resulting nucleic acids were eluted with 5 ml RNase-free deionized water and further concentrated using ethanol precipitation overnight and re-eluted in 100 ml of TE buffer. The resulting nucleic acids were then further processed for DNA and RNA separation and purification. Briefly, the extracted nucleic acids were aliquoted into two 2 ml tubes at a ratio of 1:2. RNase treatment and DNase treatment (Roche) were performed following manufacturer instructions on the recovery of DNA and RNA. After the treatments, DNA and RNA were further purified by phenol:chloroform purification to remove the enzymes and impurities. DNA and RNA were then ethanol precipitated and the pellets were eluted in TE buffer. Final purified DNA and RNA were quantified with Qubit 3.0. Quality of extracted RNA was also examined using TapeStation (Agilent) analysis at the Genome Shared Resources at the Ohio State University. Extracts were stored at –80 °C until downstream sequencing analysis.

Metatranscriptome analysis

Metatranscriptome libraries were prepared for the 27 July 2016 samples. Using 10 ng of RNA as input, ribosomal RNA was depleted using the QIAseq FastSelect –5S/16S/23S (QIAGEN, 335921) kit following the kit protocol with the following modifications: addition of probes for plant and yeast, and using one-third of the probe volumes. Then, the TruSeq Stranded mRNA Library Preparation kit (Illumina, 20020595) was used to prepare the sequencing library. Libraries were sequenced on an Illumina NovaSeq 6000 system (v.1.5 chemistry, S4 flow cell, 2 × 150 bp) at the Genomics Shared Resource Facility at the University of Colorado Anschutz Medical Campus.

Raw metatranscriptome reads were quality trimmed and adapters removed using bbdut⁵⁹ with the following flags: k = 23 mink=11 hdist=1

qtrim=rl trimq=20 minlength=75. Reads were filtered using rqcfilter2 (ref. 59) with the following flags: jni=t rna=t trimfragadapter=t qtrim=r trimq=0 maxns=1 maq=10 minlen=51 mlf=0.33 phix=t removeribo=t removehuman=t removedog=t removecat=t removemouse=t khist=t removemicrobes=t mtst=t sketch kapa=t clumpify=t tmpdir=null barcodefilter=f trimpolyg=5. Filtered and trimmed reads were mapped against the database of 97% dereplicated MAGs using bowtie2 (ref. 56) with the following flags: -D 10 -R 2 -N 1 -L 22 -i S,0,2,50. The output SAM file was converted to BAM using samtools⁵⁷ and filtered using the reformat.sh script in the bbtools package using: idfilter=0.95 pairedonly=t primaryonly=t. Mapped reads were counted using htseq-count (v.0.13.5)⁶⁰ with the following flags: -a 0 -t CDS -i ID -stranded=reverse. Read counts were filtered to remove counts <5 and were converted to geTMM⁶¹ in R. Scripts for processing metatranscriptome count data are provided in GitHub (see Data Availability). Metatranscriptome data are given in Supplementary Data 3 and gene-level mapping data are deposited in Zenodo⁶².

CAMPER polyphenol annotation construction and curation

CAMPER uses a mixture of database-derived and custom annotation searches. Database-derived searches include 272 Hidden-Markov Models (HMMs) from KofamScan⁶³, using the provided bitscore cut-offs for each HMM, and 5 HMMs from dbCAN2 (ref. 64) using the suggested bitscore cut-off (bitscore=105). For the custom annotation searches, we constructed 8 HMMs and 33 Basic Local Alignment Search Tool (BLAST) searches (Supplementary Table 1). HMMs were created in cases where sufficient biochemically characterized sequences existed (defined as sequences from at least 3 microbial genera, see Supplementary Table 1). We refer to these biochemically characterized sequences as 'seed sequences'. An overview of the HMM construction workflow is given in Supplementary Fig. 13. Briefly, for each gene type, we identified 'decoy sequences' that represented related but functionally distinct sequences to the gene of interest. To create the HMM profiles, we used BLASTP to compare the seed sequences and the decoy sequences against the UniProt90 (ref. 65) database and pulled the top 200 hits for each gene (referred to here as seed homologues and decoy homologues, respectively). Using the seed and decoy homologue sequences, seed sequences and decoy sequences, we created a sequence alignment using MAFFT (v.7.055b)⁶⁶ with '-auto' flag, and the alignment was trimmed using trimal (v.1.4.rev22) -gappypout⁶⁷. This trimmed alignment was fed to IQTree⁶⁸ (v.1.6.8) with the following flags: -alrt 1000 -bb 1000 -m MFP -nt AUTO -ntmax 10. The 'tree' file was visualized in iTOL⁶⁹ and rooted on the clade containing all decoy sequences. Then, we identified the clade that contained all seed sequences and pulled the seed homologue sequences from this clade. Importantly, we removed the seed sequences from this set to serve as a hold-out dataset when evaluating HMM quality. Using the seed homologue sequences, we constructed HMMs in graftM⁷⁰ (v.0.13.0) using the graftM create command, and searched them against the seed sequences, seed homologue sequences, decoy homologues and decoy sequences using the graftM graft command. Seed sequences, trimmed alignments, tree files and HMM seed alignments can be found on the CAMPER GitHub page (<https://github.com/WrightonLabCSU/CAMPER/>).

We curated score cut-offs to balance annotation precision and recall. We made two score cut-offs: the 'A' score was the lowest bitscore assigned to a seed sequence. As these sequences were not used in HMM construction, this score was designed to increase annotation precision (Supplementary Fig. 14). The 'B' score was curated from a combination of tree placement and score distribution. We provided a B-score to increase annotation recall (Supplementary Fig. 14), recognizing the potential for organismal bias in the seed sequences. If there was not a clear 'B' score cut-off, we opted to only provide an 'A' score. Importantly, we ensured that 'A' and 'B' score thresholds were greater than bitscores obtained from searching the HMM against decoy sequences and homologues to minimize false positive annotations. BLAST searches were

used for genes lacking sufficient numbers of and/or diversity in seed sequences (Supplementary Table 1). We used BLAST bitscores of 200 and 120 as the 'A' and 'B' score thresholds, respectively.

We validated CAMPER on a set of 5 experimentally characterized isolate genomes (GCA_004345805, GCA_000469345, GCA_000409755, GCA_000478885, GCA_003725955)⁷¹. We show that CAMPER genome annotations match the experimental phenotypes for these isolates across three polyphenol metabolic pathways (Supplementary Fig. 14 and Data 2). These metabolisms were covered by 7 HMMs and 1 BLAST search. Owing to the lack of experimental data paired to genome sequences, validation of the remaining 1 HMM and 32 BLAST searches was not possible. We note that similar to all homology-based gene annotators, CAMPER provides predicted annotations. Users should carry out additional curation, when possible; for example, curating pathway completion or checking gene organization (known gene organization is noted in the CAMPER output 'note' column).

CAMPER was used to annotate the 97% dereplicated MAG set, and all A and B-score annotations were noted. These annotations were combined with DRAM annotation outputs to evaluate the 100 transformation pathways shown in Supplementary Data 2. MAGs were said to encode a transformation pathway if they possessed >50% of the genes in the pathway. In the genome-resolved metatranscriptome data, MAGs were said to express the genes of a pathway if they encoded >50% of the genes in the pathway, and at least 1 gene recruited metatranscriptome reads (read counts >5).

We defined two metrics of polyphenol importance at the genus level: talent and dominance. To assess this, we aggregated the expression of MAGs within each genus. Polyphenol talent was defined at the genus level as expression of genes involved in at least 15 unique transformation pathways in each habitat (across depths). This was chosen as 15 pathways corresponded to a *P* value of 0.05 in the overall distribution (Supplementary Fig. 10). Polyphenol dominance was defined at the genus level as contributing an average of at least 10% of polymer, monomer or phenol active expression in a habitat (per depth). This was chosen as 10% corresponded to a *P* value of 0.05 in the overall distribution (Supplementary Fig. 10).

MAG phylogenetic tree construction

Phylogenetic trees of the MAGs were inferred using the GTDB de_novo_wf workflow for both bacterial and archaeal MAGs, using p_Patescibacteria and p_Micrarchaeota as outgroups, respectively. Phylogenetic trees were visualized using the ggtreeEXTRA^{72,73} R package.

Methanogen and methanotroph annotation curation

To confirm the identity of environmental methanogen 1,4-benzoquinone reductase (AA6) and quercetin-2,3-dioxygenase (K06911) sequences, all were submitted to Phyre2 (ref. 74) for structural modelling to confirm proposed functions. Following modelling, further validation was performed via alignment of methanogen sequences to biochemically validated or crystallized proteins using Clustal⁷⁵ to look for conservation of catalytic residues. The peroxidase sequences were also analysed via the Peroxiscan tool of RedOxiBase⁷⁶ to confirm family placement, and both the peroxidases and the quercetin-2,3-dioxygenases were analysed using SignalP⁷⁷ and PSORTb⁷⁸ to confirm their cytoplasmic status. Details on this curation can be found in Supplementary Data 5.

Enzyme assays

The activities of the five hydrolytic enzymes β -D-glucosidase, β -D-xylosidase, *N*-acetyl- β -D-glucosaminidase, arylsulphatase and phosphatase were assessed using fluorometric enzyme assays following methods adapted from refs. 79,80. Briefly, a soil slurry was prepared for each sample by blending 1 g of peat with 125 ml of sodium acetate buffer (50 mM, pH 6.2). The soil slurry was then transferred to a 96-well flat-bottom black microplate which included a buffer-only control column, as well as controls containing only soil

and standard. Subsequently, 4-methylumbelliferyl (MUB) standard solution and fluorescently linked enzyme substrates were added to the respective wells, and the plates were incubated at 25 °C for 45 min for β -D-glucosidase, *N*-acetyl- β -D-glucosaminidase, arylsulphatase and phosphatase, and for 30 min for β -D-xylosidase. Incubation times and substrate concentrations were chosen on the basis of a *V*-max test performed to capture peak enzyme activity. Fluorescence was read in a BioTek Synergy HT microplate reader at a wavelength of 460 nm emission and 360 nm excitation. Final enzyme activity was reported as $\mu\text{mol activity g}^{-1} \text{ dry soil h}^{-1}$.

The oxidative enzyme activity of phenol oxidase was assessed using a colorimetric enzyme assay⁸⁰. A soil slurry was prepared for each sample by blending 1 g of peat with 125 ml of sodium acetate buffer (50 mM, pH 6.2), and slurries were transferred to a 96-deep-well plate. A blank column containing only buffer was included in the plate, as well as controls containing only buffer and substrate. L-DOPA was chosen as the substrate for measuring polyphenol oxidase activity. After adding 25 mM L-DOPA substrate, plates were incubated at 25 °C for 24 h. Following incubation, the supernatant was transferred to a 96-well flat-bottom clear microplate and absorbance was read in a BioTek Synergy HT microplate reader at 460 nm. Final activity was reported as $\mu\text{mol activity g}^{-1} \text{ dry soil h}^{-1}$. Enzyme assay data are provided in Supplementary Data 1, and are available in Zenodo⁸¹.

Folin–Ciocalteu assay

To determine total phenolics, we used the Folin–Ciocalteu assay on the water extracts of the soil samples. Extracts were prepared from all the samples, centrifuged to remove soil and stored at –80 °C until analysis. On the day of analysis, each extract was thawed and centrifuged briefly (10,000 $\times g$) to remove insoluble material, and 25 μl aliquots were transferred to 96-well plates. The methyl-gallate standard (0.2 mg ml^{–1}) was dispensed into the wells to obtain a series of samples containing 0–4.25 μg methyl gallate. On each plate, both samples and standards were run in triplicate. Each extract was analysed on two different days. After adding 50 μl of Folin reagent (Sigma), 20 μl of 20% (m/v) Na₂CO₃ and enough water to bring the volume of each well to 95 μl , the plate was briefly mixed and then incubated in the dark for 20 min. The samples were then read at 750 nm (Biotek Synergy LX), a standard curve was generated and the total phenolic content of each sample established by interpolation. Folin–Ciocalteu assay data are provided in Supplementary Data 1.

Organic matter metabolite extraction

Water-soluble metabolites were extracted from peat by adding 7 ml of autoclaved milliQ water to 1 g of wet peat in a sterile 15 ml centrifuge tube. Tubes were vortexed twice for 30 s, and then the peat–water mixture was sonicated for 2 h at 22 °C. Samples were then centrifuged to separate the supernatant (6 ml) which served as the water extract. Samples were kept at –80 °C until analysis.

LC–MS

Samples were prepared for LC–MS/MS metabolomics as previously described⁵⁰. Briefly, water-extracted metabolites were thawed at room temperature and centrifuged again to remove any particles that potentially formed after thawing. Next, each sample was split into two 2 ml glass tube vials (1 ml each), one for hydrophilic interaction liquid chromatography (HILIC) and the other for reverse-phase liquid chromatography (RPLC). Samples in both vials were dried down completely on a Vacufuge plus (Eppendorf) and resuspended in a solution of 50% acetonitrile and 50% water for HILIC and a solution of 80% water and 20% HPLC-grade methanol for RPLC.

A Thermo Scientific Vanquish Duo ultra-high performance liquid chromatography system (UHPLC) was used for the liquid chromatography step. Extracts were separated using a Waters ACQUITY HSS T3 C18

column for RP separation and a Waters ACQUITY BEH amide column for HILIC separation.

Samples were injected in a 1 μl volume on the column and eluted as follows: for RPLC, the gradient went from 99% mobile phase A (0.1% formic acid in H₂O) to 95% mobile phase B (0.1% formic acid in methanol) over 16 min. For HILIC, the gradient went from 99% mobile phase A (0.1% formic acid, 10 mM ammonium acetate, 90% acetonitrile, 10% H₂O) to 95% mobile phase B (0.1% formic acid, 10 mM ammonium acetate, 50% acetonitrile, 50% H₂O). Both columns were run at 45 °C at a flow rate of 300 $\mu\text{l min}^{-1}$.

A Thermo Scientific Orbitrap Exploris 480 was used for spectral data collection at a spray voltage of 3,500 V for positive mode (for RPLC) and 2,500 V for negative mode (for HILIC) using the heated-electrospray ionization (H-ESI) source. The ion transfer tube and vaporizer temperature were both 350 °C. Compounds were fragmented using data-dependent MS/MS with HCD collision energies of 20, 40 and 80.

The Compound Discoverer 3.3.2.31 software (Thermo Fisher) was used to analyse the data using the untargeted metabolomics workflow. Briefly, the spectra were first aligned, followed by a peak-picking step. Putative elemental compositions of unknown compounds were predicted using the exact mass, isotopic pattern, fine isotopic pattern and MS/MS data using the built-in HighChem Fragmentation Library of reference fragmentation mechanisms. Metabolite annotation was performed using an in-house database built using 1,200 reference standards, spectral libraries and compound databases. First, fragmentation scans, retention times and ion masses of unknown compounds were compared with those in the in-house database. Second, fragmentation scans (MS²) searches were performed in mzCloud, which is a curated database of MSⁿ spectra containing more than 9 million spectra and 20,000 compounds. Third, predicted compositions were obtained on the basis of mass error, matched isotopes, missing number of matched fragments, spectral similarity score (calculated by matching theoretical and measured isotope patterns), matched intensity percentage of the theoretical pattern, the relevant portion of MS and the MS/MS scan. The mass tolerance used for estimating predicted composition was 5 ppm. Finally, annotation was complemented by searching MS¹ scans in different online databases with ChemSpider (using either the exact mass or the predicted formula). The compounds' level of annotation was assigned according to the Metabolomics Standards Initiative as follows⁸²: Level 1: compounds with exact match to a standard reference compound in our in-house library; Level 2: compounds with full match to online spectral databases using the mzCloud database (based on MS² spectra matching) and ChemSpider (using mass and molecular formula generated through the Predicted Compositions node). Annotations were manually inspected for validation, identification of isomers and removal of in-source fragments.

To enhance annotation coverage and add compound classes, SIRIUS, CSI:FingerID and CANOPUS were used⁸³. Compound chemical taxonomy based on chemical structure was assigned using ClassyFire⁸⁴. For annotated polyphenols, chemical taxonomy was assigned by hand using the following rules: phenolic acids contain a single aromatic ring with only hydroxyl and ether groups attached; benzoic acids contain a single aromatic ring with a carboxylic acid or ester attached in the alpha position of the carbonyl; flavonoids are 15-carbon members of a heteroatom connected to an aromatic ring; lignans contain two aromatic rings bound by a 4-carbon chain; stilbenes are two aromatic rings separated by two carbons with a double bond between them; and others are aromatic compounds that do not fit into the above categories. LC–MS/MS data are given in Supplementary Data 4.

Fourier-transform ion cyclotron resonance MS analysis

Water extracts (3 ml) were first purified using solid-phase extraction to remove contaminants (salts) according to ref. ⁸⁵. Briefly, water extracts were acidified to pH 2 using 1 M HCl. Then, extracts were filtered through

a 3 ml Bond Elut PPE cartridge (Agilent) that was previously activated using methanol. Cartridges were washed with 3 ml of a 0.01 M HCl solution five times, then dried using filtered air. Finally, extracts were eluted using 1.5 ml of methanol and stored at -80°C until use.

Purified extracts were analysed by direct injection using a 12 Tesla Bruker Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer located at the Pacific Northwest National Laboratory. Negative charged molecular ions were generated using a Bruker ESI source. Instrument stability was optimized using a Suwannee River Fulvic Acid standard (SRFA), obtained from the International Humic Substance Society. Potential carry-over between samples was monitored by injecting HPLC-grade methanol. The instrument was flushed between samples using a combination of milliQ water and methanol. To account for variations in carbon concentrations in different samples, the ion accumulation time was varied between 0.03 and 0.05 s. A total of 144 individual scans per sample were collected, averaged and calibrated using an organic matter homologous series separated by 14 Da (CH₂). Mass accuracy was <1 ppm for single charged ions measured across an m/z range of 100–1,200 m/z , the mass resolution was $\sim 240\text{ K}$ at 341 m/z and the transient was 0.8 s. Raw spectra collected per sample were transformed into a list of m/z values using the FT-MS peak picker module within the Bruker Daltonik (v.4.2) software using a signal to noise ratio of 7 and absolute intensity threshold of 100 (default). Formularity⁸⁶ software was used to assign putative chemical formulae following ref. 87. FT-ICRMS data are available in Supplementary Data 4 and in Zenodo⁸⁸.

Refinement of tannin and lignin Van Krevlen classes

Van Krevlen analyses assign molecular formulae to classes using H:C and O:C ratios⁸⁹. Two classical regions are the lignin-like and tannin-like regions; however, these regions mask a huge amount of complexity (for example, condensed tannins vs hydrolysable tannins). Using an approach similar to that in ref. 90, we sought to refine these classifications using known formulae from over 60 characterized natural substrates (Supplementary Data 6 and Fig. 15)^{91–102}. We determined the following H:C and O:C ratios for hydrolysable tannins (HT), condensed tannins (CT) and lignin (lig):

Lignin: $0.3 < \text{O:C} < 0.48$, $1.08 < \text{H:C} < 1.28$

Hydrolysable tannin: $0.6 < \text{O:C} < 0.7$, $0.58 < \text{H:C} < 0.89$

Condensed tannin: $0.4 < \text{O:C} < 0.5$, $0.74 < \text{H:C} < 0.88$

We used the CT, HT and lig class boundaries to classify features. To determine the relative abundance of these classes, we summed the abundance of features for each class within each sample. We then averaged this across replicates for each class within a habitat and depth. To determine ‘percent polyphenol’ (Fig. 1), we summed the percent CT, percent HT and percent lig.

Statistics

All data analyses and visualization were done in R (v.4.2.1)¹⁰³ with the following packages: stats, cowplot (v.1.1.1)¹⁰⁴, ggplot2 (v.3.3.6)¹⁰⁵, ggtree-EXTRA (v.1.6.0)⁷², tidyr (v.1.2.0)¹⁰⁶, dplyr (v.1.0.9)¹⁰⁷, readxl (v.1.4.0)¹⁰⁸, pheatmap (v.1.0.12)¹⁰⁹, RColorBrewer (v.1.1-3)¹¹⁰, edgeR (v.3.16)¹¹¹, stringr (v.1.4.0)¹¹² and writexl (v.1.4.0)¹¹³. For statistical analyses of enzyme latch data types, we used Mann–Whitney test and Pearson correlations in R, and P values were adjusted within groups of comparisons using Benjamini–Hochberg adjustment.

Reporting summary

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

Data availability

The metagenomes, metatranscriptomes and most metagenome-assembled genomes used in this paper are available at NCBI under

BioProjectID PRJNA386568. See Supplementary Data 3 for individual BioSample numbers. All MAGs are available via Zenodo at <https://doi.org/10.5281/zenodo.7596016> (ref. 114). All raw and processed data are available in the following Zenodo archives: MAG annotations at <https://doi.org/10.5281/zenodo.7587534> (ref. 55), metatranscriptome mapping at <https://doi.org/10.5281/zenodo.7591900> (ref. 62), FT-ICRMS at <https://doi.org/10.5281/zenodo.7519321> (ref. 88) and enzyme assays at <https://doi.org/10.5281/zenodo.7519395> (ref. 81). The following publicly available databases were used in this study: UniProt90 (<https://www.uniprot.org/help/uniref>), RedOxiBase (<https://peroxibase.toulouse.inra.fr/>), SignalP (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>), pSortB (<https://www.psort.org/psortb/>) and KOfam (<https://www.genome.jp/ftp/db/kofam/>).

Code availability

All scripts involved with data generation, processing, curation and visualization are available on GitHub (https://github.com/WrightonLabCSU/stordalen_polyphenols_paper). The CAMPER annotation tool is available on GitHub (<https://github.com/WrightonLabCSU/CAMPER>).

References

- Hugelius, G. et al. Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **11**, 6573–6593 (2014).
- Schuur, E. A. G. et al. Permafrost and climate change: carbon cycle feedbacks from the warming arctic. *Annu. Rev. Environ. Resour.* **47**, 343–371 (2022).
- Natali, S. M. et al. Permafrost carbon feedbacks threaten global climate goals. *Proc. Natl Acad. Sci. USA* **118**, e2100163118 (2021).
- Turetsky, M. R. et al. Carbon release through abrupt permafrost thaw. *Nat. Geosci.* **13**, 138–143 (2020).
- Woodcroft, B. J. et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
- Koven, C. D. et al. Permafrost carbon–climate feedbacks accelerate global warming. *Proc. Natl Acad. Sci. USA* **108**, 14769–14774 (2011).
- Freeman, C., Ostle, N. & Kang, H. An enzymic ‘latch’ on a global carbon store: a shortage of oxygen locks up carbon in peatlands by restraining a single enzyme. *Nature* **409**, 149 (2001).
- Fenner, N. & Freeman, C. Drought-induced carbon loss in peatlands. *Nat. Geosci.* **4**, 895–900 (2011).
- Fenner, N. & Freeman, C. Woody litter protects peat carbon stocks during drought. *Nat. Clim. Change* **10**, 363–369 (2020).
- Zak, D. et al. Unraveling the importance of polyphenols for microbial carbon mineralization in rewetted riparian peatlands. *Front. Environ. Sci.* <https://doi.org/10.3389/fenvs.2019.00147> (2019).
- Cheyrier, V. Polyphenols in foods are more complex than often thought. *Am. J. Clin. Nutr.* **81**, 223S–229S (2005).
- Urbanová, Z. & Hájek, T. Revisiting the concept of ‘enzymic latch’ on carbon in peatlands. *Sci. Total Environ.* **779**, 146384 (2021).
- Hall, S. J., Treffkorn, J. & Silver, W. L. Breaking the enzymatic latch: impacts of reducing conditions on hydrolytic enzyme activity in tropical forest soils. *Ecology* **95**, 2936–2945 (2014).
- Brouns, K., Keuskamp, J. A., Potkamp, G., Verhoeven, J. T. A. & Hefting, M. M. Peat origin and land use effects on microbial activity, respiration dynamics and exo-enzyme activities in drained peat soils in the Netherlands. *Soil Biol. Biochem.* **95**, 144–155 (2016).
- Romanowicz, K. J. et al. Understanding drivers of peatland extracellular enzyme activity in the PEATcosm experiment: mixed evidence for enzymic latch hypothesis. *Plant Soil* **397**, 371–386 (2015).
- Wen, Y. et al. Is the ‘enzyme latch’ or ‘iron gate’ the key to protecting soil organic carbon in peatlands? *Geoderma* **349**, 107–113 (2019).

17. Renault, H. et al. A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat. Commun.* **8**, 14713 (2017).
18. Berland, H. et al. Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants. *Proc. Natl Acad. Sci. USA* **116**, 20232–20239 (2019).
19. Scalbert, A. Antimicrobial properties of tannins. *Phytochemistry* **30**, 3875–3883 (1991).
20. Walter, K. & Schütt, C. in *Methods of Enzymatic Analysis* 2nd edn (ed. Bergmeyer, H. U.) 856–860 (Academic Press, 1974).
21. Drula, E. et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).
22. Dennis, D. & Kaplan, N. O. D- and L-lactic acid dehydrogenases in *Lactobacillus plantarum*. *J. Biol. Chem.* **235**, 810–818 (1960).
23. Goris, T., Cuadrat, R. R. C. & Braune, A. Flavonoid-modifying capabilities of the human gut microbiome—an in silico study. *Nutrients* **13**, 2688 (2021).
24. Johansson, T. et al. Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Global Change Biol.* **12**, 2352–2369 (2006).
25. Fofana, A. et al. Mapping substrate use across a permafrost thaw gradient. *Soil Biol. Biochem.* **175**, 108809 (2022).
26. Cory, A. B. et al. Quantifying the inhibitory impact of soluble phenolics on anaerobic carbon mineralization in a thawing permafrost peatland. *PLoS ONE* **17**, e0252743 (2022).
27. Hodgkins, S. B. et al. Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. *Proc. Natl Acad. Sci. USA* **111**, 5819–5824 (2014).
28. Dilling, S., Imkamp, F., Schmidt, S. & Müller, V. Regulation of caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*. *Appl. Environ. Microbiol.* **73**, 3630–3636 (2007).
29. McGivern, B. B., Flynn, R. M., Woyda, R. & Wrighton, K. C. CAMPER: Curated Annotations for Microbial Polyphenol Enzymes and Reactions. *Zenodo* <https://doi.org/10.5281/zenodo.10835273> (2023).
30. Rothwell, J. A. et al. Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013**, bat070 (2013).
31. Hough, M. et al. Coupling plant litter quantity to a novel metric for litter quality explains C storage changes in a thawing permafrost peatland. *Global Change Biol.* **28**, 950–968 (2022).
32. McGivern, B. B. et al. Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat. Commun.* **12**, 2466 (2021).
33. DeAngelis, K. M. et al. Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. *Front. Microbiol.* **4**, 280 (2013).
34. Chirania, P. et al. Metaproteomics reveals enzymatic strategies deployed by anaerobic microbiomes to maintain lignocellulose deconstruction at high solids. *Nat. Commun.* **13**, 3870 (2022).
35. Hess, V., Gonzalez, J. M., Parthasarathy, A., Buckel, W. & Muller, V. Caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*: a coenzyme A loop saves energy for caffeate activation. *Appl. Environ. Microbiol.* **79**, 1942–1947 (2013).
36. Perez, J. M. et al. Redundancy in aromatic O-demethylation and ring-opening reactions in *Novosphingobium aromaticivorans* and their impact in the metabolism of plant-derived phenolics. *Appl. Environ. Microbiol.* **87**, e02794–20 (2021).
37. Kolton, M. et al. Defining the sphagnum core microbiome across the North American continent reveals a central role for diazotrophic methanotrophs in the nitrogen and carbon cycles of boreal peatland ecosystems. *mBio* **13**, e03714–e03721 (2022).
38. Hausmann, B. et al. Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME J.* **12**, 1729–1742 (2018).
39. Tveit, A. T., Urich, T., Frenzel, P. & Svenning, M. M. Metabolic and trophic interactions modulate methane production by Arctic peat microbiota in response to warming. *Proc. Natl Acad. Sci. USA* **112**, E2507–E2516 (2015).
40. Mondav, R. et al. Discovery of a novel methanogen prevalent in thawing permafrost. *Nat. Commun.* **5**, 3212 (2014).
41. Field, J. A., Kortekaas, S. & Lettinga, G. The tannin theory of methanogenic toxicity. *Biol. Wastes* **29**, 241–262 (1989).
42. Bowater, L., Fairhurst, S. A., Just, V. J. & Bornemann, S. *Bacillus subtilis* YxaG is a novel Fe-containing quercetin 2,3-dioxygenase. *FEBS Lett.* **557**, 45–48 (2004).
43. El Hadrami, A., Islam, Md. R., Adam, L. R. & Daayf, F. A cupin domain-containing protein with a quercetinase activity (VdQase) regulates *Verticillium dahliae*'s pathogenicity and contributes to counteracting host defenses. *Front. Plant Sci.* **6**, 440 (2015).
44. Daniels, L., Fuchs, G., Thauer, R. K. & Zeikus, J. G. Carbon monoxide oxidation by methanogenic bacteria. *J. Bacteriol.* **132**, 118–126 (1977).
45. Ferry, J. G. CO in methanogenesis. *Ann. Microbiol.* **60**, 1–12 (2010).
46. Dunfield, P. F. & Dedysh, S. N. *Methylocella*: a gourmand among methanotrophs. *Trends Microbiol.* **22**, 368–369 (2014).
47. Belova, S. E. et al. Acetate utilization as a survival strategy of peat-inhabiting *Methylocystis* spp. *Environ. Microbiol. Rep.* **3**, 36–46 (2011).
48. Lankiewicz, T. S. et al. Lignin deconstruction by anaerobic fungi. *Nat. Microbiol.* **8**, 596–610 (2023).
49. Janusz, G. et al. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol. Rev.* **049**, 941–962 (2017).
50. Ellenbogen, J. B. et al. Methylophony in the Mire: direct and indirect routes for methane production in thawing permafrost. *mSystems* **9**, e00698-23 (2023).
51. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
52. Woodcroft, B. J. Galah. *GitHub* <https://github.com/wwood/galah> (2022).
53. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927 (2020).
54. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).
55. McGivern, B., Cronin, D. & NSF EMERGE Biology Integration Institute. Metagenome-assembled genome DRAM annotations (EMERGE 97% dereplicated MAGs). *Zenodo* <https://doi.org/10.5281/zenodo.7587534> (2023).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Woodcroft, B. J. CoverM: read coverage calculator for metagenomics. *GitHub* <https://github.com/wwood/CoverM> (2020).
59. Bushnell, B. BBtools <https://sourceforge.net/projects/bbmap/> (Joint Genome Institute, 2021).
60. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
61. Smid, M. et al. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics* **19**, 236 (2018).

62. McGivern, B. Stordalen Mire July 2016 metatranscriptome data from McGivern et al. *Zenodo* <https://doi.org/10.5281/zenodo.7591900> (2023).
63. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
64. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
65. Supek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
66. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
67. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
68. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
69. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
70. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* **46**, e59 (2018).
71. Braune, A. & Blaut, M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* **7**, 216–234 (2016).
72. Xu, S. et al. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Mol. Biol. Evol.* **38**, 4039–4042 (2021).
73. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
74. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
75. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).
76. Savelli, B. et al. RedoxiBase: a database for ROS homeostasis regulated proteins. *Redox Biol.* **26**, 101247 (2019).
77. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
78. Gardy, J. L. et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21**, 617–623 (2005).
79. Saiya-Cork, K. R., Sinsabaugh, R. L. & Zak, D. R. The effects of long term nitrogen deposition on extracellular enzyme activity in an *Acer saccharum* forest soil. *Soil Biol. Biochem.* **34**, 1309–1315 (2002).
80. DeForest, J. L. The influence of time, storage temperature, and substrate age on potential soil enzyme activity in acidic forest soils using MUB-linked substrates and L-DOPA. *Soil Biol. Biochem.* **41**, 1180–1186 (2009).
81. Hernandez, A., Bernhardt, L., Daly, R., McGivern, B. & Ernakovich, J. EMERGE 2016 autochamber sites enzyme assays (O.O.O). *Zenodo* <https://doi.org/10.5281/zenodo.7519395> (2023).
82. Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
83. Dührkop, K. et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **16**, 299–302 (2019).
84. Djoumbou Feunang, Y. et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **8**, 61 (2016).
85. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **6**, 230–235 (2008).
86. Tolić, N. et al. Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Anal. Chem.* **89**, 12659–12665 (2017).
87. Tfaily, M. M. et al. Vertical stratification of peat pore water dissolved organic matter composition in a peat bog in northern Minnesota. *J. Geophys. Res. Biogeosci.* **123**, 479–494 (2018).
88. Freire Zapata, V., McGivern, B., Daly, R. & Tfaily, M. EMERGE 2016 autochamber sites FT-ICRMS (O.O.O). *Zenodo* <https://doi.org/10.5281/zenodo.7519321> (2023).
89. Quideau, S., Deffieux, D., Douat-Casassus, C. & Pouységu, L. Plant polyphenols: chemical properties, biological activities, and synthesis. *Angew. Chem. Int. Ed.* **50**, 586–621 (2011).
90. Laszakovits, J. R. & MacKay, A. A. Data-based chemical class regions for Van Krevelen diagrams. *J. Am. Soc. Mass Spectrom.* **33**, 198–202 (2022).
91. Reeves, S. G. et al. Proanthocyanidin structural details revealed by ultrahigh resolution FT-ICR MALDI-Mass Spectrometry, 1H–13C HSQC NMR, and Thiolytic-HPLC–DAD. *J. Agric. Food Chem.* **68**, 14038–14048 (2020).
92. Li, C. et al. High molecular weight persimmon (*Diospyros kaki* L.) proanthocyanidin: a highly galloylated, A-linked tannin with an unusual flavonol terminal unit myricetin. *J. Agric. Food Chem.* **58**, 9033–9042 (2010).
93. Yokota, K., Kimura, H., Ogawa, S. & Akihiro, T. Analysis of A-type and B-type highly polymeric proanthocyanidins and their biological activities as nutraceuticals. *J. Chem.* **2013**, e352042 (2012).
94. Naumann, H. et al. Relationships between structures of condensed tannins from Texas legumes and methane production during in vitro rumen digestion. *Molecules* **23**, 2123 (2018).
95. Sarni-Manchado, P., Cheynier, V. & Moutounet, M. Interactions of grape seed tannins with salivary proteins. *J. Agric. Food Chem.* **47**, 42–47 (1999).
96. Scioneaux, A. N. et al. Qualitative variation in proanthocyanidin composition of *Populus* species and hybrids: genetics is the key. *J. Chem. Ecol.* **37**, 57–70 (2011).
97. Bar-Ya'akov, I., Tian, L., Amir, R. & Holland, D. Primary metabolites, anthocyanins, and hydrolyzable tannins in the pomegranate fruit. *Front. Plant Sci.* **10**, 620 (2019).
98. Engström, M. T. et al. The oxidative activity of ellagitannins dictates their tendency to form highly stabilized complexes with bovine serum albumin at increased pH. *J. Agric. Food Chem.* **64**, 8994–9003 (2016).
99. Ito, H. et al. Dimeric and trimeric hydrolyzable tannins from *Quercus coccifera* and *Quercus suber*. *J. Nat. Prod.* **65**, 339–345 (2002).
100. Ito, H., Miki, K. & Yoshida, T. Elaeagnatins A-G, C-glucosidic ellagitannins from *Elaeagnus umbellata*. *Chem. Pharm. Bull.* **47**, 536–542 (1999).
101. Moilanen, J. & Salminen, J.-P. Ecologically neglected tannins and their biologically relevant activity: chemical structures of plant ellagitannins reveal their in vitro oxidative activity at high pH. *Chemoecology* **18**, 73–83 (2008).

102. Moilanen, J. et al. Biological activity of ellagitannins: effects as anti-oxidants, pro-oxidants and metal chelators. *Phytochemistry* **125**, 65–72 (2016).
103. R Foundation. *The R Project for Statistical Computing* <https://www.r-project.org/> (2022).
104. Wilke, C. O. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. R package v.1.1.3 <https://wilkelab.org/cowplot/> (2020).
105. Wickham, H. *Ggplot2. Elegant Graphics for Data Analysis* (Springer, 2016).
106. Wickham, H., Vaughan, D. & Girlich, M. tidy: tidy messy data. R package v.1.3.1 <https://tidyr.tidyverse.org/> (2024).
107. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: a grammar of data manipulation. R package v.1.1.4 <https://dplyr.tidyverse.org> (2023).
108. Wickham, H. & Bryan, J. readxl: read Excel files. *GitHub* <https://github.com/tidyverse/readxl> (2023).
109. Kolde, R. pheatmap: Pretty Heatmaps <https://cran.r-project.org/package=pheatmap> (CRAN, 2019).
110. Neuwirth, E. RColorBrewer: ColorBrewer Palettes (CRAN, 2022).
111. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
112. Wickham, H. stringr: simple, consistent wrappers for common string operations. *GitHub* <https://github.com/tidyverse/stringr> (2022).
113. Ooms, J. & McNamara, J. writexl: export data frames to Excel 'xlsx' format. R package v.1.5.0 <https://docs.ropensci.org/writexl/> (2023).
114. Cronin, D. & NSF EMERGE Biology Integration Institute. Metagenome-assembled genomes (MAGs) from Stordalen Mire, Sweden (O.O.O-beta_v2). *Zenodo* <https://doi.org/10.5281/zenodo.7596016> (2023).

Acknowledgements

We thank the EMERGE Biology Integration Institute Coordinators (members listed in Supplementary Information) for project guidance and management, the EMERGE 2016 Field and Analysis Team (members listed in Supplementary Information) for field work and sample processing, M. Shaffer and A. Narowe for methods consultations, and R. Daly for project coordination. B.B.M. and K.C.W. were fully or partially supported by an Early Career Award to K.C.W. from the National Science Foundation (NSF) under award number 1912915. K.C.W., J.B.E., M.A.B., R.W. and R.M.F. were fully or partially supported by the US Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research (BER) (grants DE-SC0019746, DE-SC0021350, DE-SC0023456 and DE-SC0023084) and NSF award number 2149505. B.J.W. was supported by an Australian Research Council Future Fellowship (no. FT210100521). This research is a contribution of the EMERGE Biology Integration Institute, which also supports K.C.W., B.B.M., D.R.C., J.B.E., V.F.-Z., L.B., A.B.C., R.M.W., J.P.C., B.J.W., J.G.E., M.M.T., M.B.S., G.W.T. and V.I.R., funded by the NSF Biology Integration Institutes Program, Award no. 2022070. This study was further supported by the Genomic Science Program of the DOE Office of BER, grants DE-SC0004632 and DE-SC0010580. Portions of this work were supported by: (1) BER Support Science Proposal 503530, conducted by the DOE Joint Genome Institute (JGI), a DOE Office of Science User Facility sponsored by the Office of BER and operated under contract no. DE-AC02-05CH11231 (JGI); (2) the Facilities Integrating Collaborations for User Science (FICUS) program (proposal: 10.46936/fics.proj.2017.49950/60006215,

10.46936/10.25585/60001148) and used resources at the DOE JGI (<https://ror.org/04xm1d337>) and the Environmental Molecular Sciences Laboratory (<https://ror.org/04rc0xn13>), which are DOE Office of Science User Facilities operated under contract nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). We also thank the Swedish Polar Research Secretariat and SITES for support of the work done at the Abisko Scientific Research Station. SITES was supported by the Swedish Research Council's grant 4.3-2021-00164. Quality control of RNA used in this publication was performed at The Ohio State University Comprehensive Cancer Center Core Facilities, supported by the Center and by the National Institutes of Health grant P30 CA016058. Metatranscriptome sequencing completed in this publication was performed at the Genomics Shared Resource at the University of Colorado Cancer Center, supported by the Cancer Center Support grant P30CA046934.

Author contributions

B.B.M. and K.C.W. conceptualized, designed and supervised the study. B.B.M., D.R.C., J.B.E., E.L.K., V.F.-Z., J.A.B., L.B., A.I.H., R.M.F., R.W., A.B.C., R.M.W., J.P.C., J.G.E., M.M.T. and A.E.H. performed and supervised experimental work. B.B.M. analysed and visualized the data. B.B.M., D.R.C. and B.J.W. curated the data. B.B.M. and K.C.W. drafted the manuscript, with contributions from M.A.B., M.B.S., G.W.T., V.I.R. and A.E.H. All authors read, commented on and edited the manuscript, and approved its final form.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-024-01691-0>.

Correspondence and requests for materials should be addressed to Kelly C. Wrighton.

Peer review information *Nature Microbiology* thanks Chris Greening, Marnix Medema and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	All scripts involved with data generation, processing, curation, and visualization are available on GitHub (https://github.com/WrightonLabCSU/stordalen_polyphenols_paper). The code for CAMPER is available on GitHub (https://github.com/WrightonLabCSU/CAMPER).
Data analysis	<p>The following published software was used in data analysis: R (v4.2.1), bbtools (v38.51), CheckM (v1.0.12), bowtie2 (v2.4.1), galah (v0.3.0), GTDB-tk (v2.0), DRAM (v1.4.0), samtools (v1.9), coverM (v0.6.0), htseq-count (v.0.13.5), bbdut (bbtools v38.51), MAFFT (v7.055b), using trimal (v1.4.rev22), IQTree (v1.6.8), graftM (v0.13.0), Compound Discover (v3.3.2.3.1), Formularity (v1.0.7947), cowplot (v1.1.1), ggplot2 (v.3.3.6), ggtreeEXTRA (v1.6.0), tidyv (v.1.2.0), dplyr (v.1.0.9), readxl (v1.4.0), pheatmap (v1.0.12), RColorBrewer (v1.1-3), edgeR (v3.16), stringr (v1.4.0), writexl (v1.4.0).</p> <p>The following databases were used for annotation within DRAM and for CAMPER: dbCAN2 (v11), KEGG/Kofam (release 104).</p>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The metagenomes, metatranscriptomes, and metagenome-assembled genomes used in this paper are available at NCBI under BioProjectID PRJNA386568. See Supplementary Table 3 for individual BioSample numbers. All raw and processed data are available in the following Zenodo archives: MAG annotations (10.5281/zenodo.7587534), metatranscriptome mapping (10.5281/zenodo.7591900), FT-ICRMS (10.5281/zenodo.7519321), and enzyme assays (10.5281/zenodo.7519395). The following publicly available databases were used in this study: UniProt90 (<https://www.uniprot.org/help/uniref>), RedOxiBase (<https://peroxibase.toulouse.inra.fr/>), SignalP (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>), pSortB (<https://www.psort.org/psortb/>), KOfam (<https://www.genome.jp/ftp/db/kofam/>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Data exclusions

Replication

Randomization

Blinding

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |