Universal microbial reworking of dissolved organic matter along environmental gradients

Erika C Freeman[1, *], Erik JS Emilson[2], Thorsten Dittmar[3,4], Lucas PP Braga[1], Caroline E Emilson[2], Tobias Goldhammer[5], Christine Martineau[6], Gabriel Singer[7], Andrew J Tanentzap[1]

[1]Ecosystems and Global Change Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[2]Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen St. E., Sault Ste. Marie, ON, P6A 2E5, Canada

[3]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, 26129 Oldenburg, Germany

[4]Helmholtz Institute for Functional Marine Biodiversity, University of Oldenburg, 26129 Oldenburg, Germany

[5]Department of Ecohydrology and Biogeochemistry, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Mueggelseedamm 301, Berlin, Germany

[6]Natural Resources Canada, Laurentian Forestry Centre, 1055 Du P.E.P.S. Street, P.O. Box 10380, Québec, G1V 4C7, Canada

[7]Department of Ecology, University of Innsbruck, Technikerstrasse 25, 6020 Innsbruck, Austria

*Author for correspondences: erika.freem@gmail.com

Abstract

Soils are losing increasing amounts of carbon annually to freshwaters as dissolved organic matter (DOM), which, if degraded, can accelerate climate change. DOM is more susceptible to degradation closer to its source and becomes increasingly dominated by the same, difficult-to-degrade compounds as degradation proceeds. However, the processes underlying DOM degradation across environments are poorly understood. Here we found DOM changed similarly along soil-aquatic gradients irrespective of differences in environmental conditions. Using ultra-high-resolution mass spectrometry, we tracked DOM along soil depths and hillslope positions in forest headwater catchments and related its composition to soil microbiomes and physical chemistry. Along depths and hillslopes, carbohydrate-like and unsaturated hydrocarbon-like compounds increased in mass, suggestive of microbial reworking of plant material. More than half of the variation in the abundance of these compounds was related to the expression of genes essential for degrading plant-derived carbohydrates. Our results implicate continuous microbial reworking in shifting DOM towards universal compounds in soils. By synthesising data from the land-to-ocean continuum, we suggest these processes can generalise across ecosystems and spatiotemporal scales. Such general degradation patterns can be leveraged to predict DOM composition and its downstream reactivity along environmental gradients to inform management of soil-to-stream carbon losses.

Introduction

The fate of carbon exported from soils into aquatic ecosystems is a poorly understood component of the global carbon cycle. Soils store at least twice as much carbon as the atmosphere[1] and are expected to absorb more than one-third of anthropogenic emissions[2]. However, >15% of the net carbon added to soils annually from decomposing plant litter and roots is leached into aquatic systems as dissolved organic matter (DOM)[3]. Once in water, much of the DOM pool is highly reactive[4], potentially returning large quantities of carbon to the atmosphere as CO_2 or CH_4 and offsetting terrestrial carbon sequestration[5].

The distance from its terrestrial sources can predict how much DOM degrades in aquatic ecosystems[6]. Terrestrial DOM is initially dominated by a few spatially heterogeneous biomolecules, such as lignin-derived polyphenols that reflect local plant species composition[7]. Continuous transformation and remineralisation of this DOM within merging flowpaths produce an increasingly homogenous pool of compounds downstream[8–10]. Compounds with structural features such as carboxylic-rich alicyclic moieties, material derived from linear terpenoids, and carotenoid degradation products dominate this pool[11]. As these compounds occur everywhere[12], that is, in all samples, they are termed “universal”. The convergence towards a DOM pool dominated by universal compounds is known as a “degradation cascade”[12]. Furthermore, the proportion of compounds in a DOM pool that is shared with the endpoint of the degradation cascade, such as measured with molecular β-diversity, should increase along this cascade[12], but this idea remains untested. The amount of time DOM is exposed to microbial and photochemical processing[13] is likely a unifying mechanism behind the degradation cascade[14]. Time influences carbon decay rates in marine sediments[15,16], bioassays[17], and inland waters[4,9]. As many universal compounds are also degraded slowly by microbes[18], the homogenised DOM pools can ultimately provide a persistent carbon store despite losses in terrestrial carbon sequestration.

Universal DOM pools can result from common synthetic pathways or a chain of similar degradation steps[6,12,19–21], but how the importance of these mechanisms changes along diverse environmental gradients remains unknown[14]. Unlike the oceans, processing of DOM during vertical soil passage does not consistently converge to low-molecular-weight, recalcitrant compounds[22]. Degradation of DOM is instead characterised by increasing weight, likely reflecting microbial transformation of small plant-derived compounds into larger microbial products[22,23]. Degradation also varies with compound concentration[24], abiotic processes like sorption and desorption to minerals[25], hydrological pathways[26], and microbial trait diversity and energy supply[27]. Thus, DOM composition, microbial metabolism, and environmental and ecosystem properties interact to stabilise carbon[28]. Understanding how these processes influence DOM degradation along different flowpaths is necessary to ensure land-based carbon sequestration efforts are not offset downstream.

Here we asked if the molecular composition of DOM changes similarly through soil depth and along hillslopes with different environmental conditions and thus potential degradation processes. We worked across four replicate headwater catchments in northwestern Ontario, Canada (Fig. S1). We advanced previous studies by focusing on soil-water flow upstream of the headwater-ocean continuum, i.e., a natural but neglected extension of the riverine continuum[29]. We tracked DOM from 5 to 60 cm soil depth at shoulder, back, toe, and foot hillslope positions and into streams using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) (Fig. 1a). We paired FT-ICR-MS with shotgun metatranscriptomic sequencing and metabolic measurements to reconstruct the function of microbial communities. Our fully factorial depth-by-hillslope design allowed us to test changes in the DOM pool originating from the same source material (i.e., within each position) but along contrasting environmental gradients. Including headwater streams also allowed us to track the fate of soil carbon downstream. Although we expected DOM to

become increasingly homogenous across both soil depth and hillslope, consistent with a degradation cascade[12], we expected the extent of homogenisation and underlying processes would differ. With depth, the DOM pool should reflect preferential consumption and transformation by microbial processes[22]. DOM sources will be relatively consistent through depth with less hydrological mixing of sources than between hillslope positions. Along the hillslope, DOM should reflect different sources because of hydrological mixing from different positions that vary in moisture, erosion, vegetation type and rooting depth, and microbial processing should be relatively less important. The absence of differences in DOM between these two spatial gradients would suggest that universal processes, such as time exposed to microbial processing, shape DOM along different degradation cascades. Our results now implicate common metabolic processes in shifting DOM towards homogenised compounds along a soil-headwater continuum and suggest this process generalises across environmental and spatiotemporal gradients.

Degradation of DOM across soil depth and hillslope

Consistent with the predicted increase in universal compounds from headwater to ocean[12], we found upland soils had fewer universal compounds than aquatic samples. We detected 9327 unique molecular formulae, more than twice that observed in a previous soil study[22], suggesting we representatively sampled the DOM pool. We attributed this result to high extraction efficiencies for DOM (mean 69% ± 6% s.d.; see Methods), technical improvements in the ICR detection cell, and advancements in molecular formula assignment[30]. Despite the many detected formulae, only 13% occurred in all samples, that is, were universal, compared with between 47 to 87% in a synthesis from headwaters to oceans[12]. Of the universal compounds, 79% were classified[31] as lignin-like based on the similarity of their elemental compositions to known biomolecules (Fig. 1b). Tannin-like (representing phenol derivatives)

and condensed hydrocarbon-like compounds were the next most abundant classes accounting for 9%, and 6% of molecular formulae, respectively. Our results were also not due to missing high-molecular-weight substances (HMWS) commonly detected in soils[32] but outside the analytical window of FT-ICR-MS[33]. Using size-exclusion chromatography, HMWS were detected in only 13% of soil pore water samples and, when found, contributed, on average (95% confidence interval, CI) only 8% (3 to 12%) to the total dissolved organic carbon pool.

DOM converged towards a universal compound pool differently between soil depth and hillslope gradients. Consistent with our predictions, universal compounds increased from an estimated mean (95% CI) of 20.9% (19.0 to 22.9%) of all formulae at 5 cm to 23.9% (21.9 to 26.1%) at 60 cm (Fig. 1c). There was no change from the shoulder position into streams (Fig. 1d; Table S1), as expected if hydrological mixing was important along the hillslope gradient. However, the DOM pool was similarly homogenised along both soil depth and hillslope gradients when we measured the proportion of signal intensity (i.e., relative abundance) attributed to universal compounds. The relative abundance of universal compounds increased by an estimated mean (95% CI) of 9.4% (5.9 to 12.9%) and 8.1% (6.4 to 9.7%) from 5 cm to 60 cm and shoulder to stream, respectively (Fig. 1e,f). Universal compounds identified as those occurring in all our samples had similar molecular properties to literature definitions of degradation end-products that were independent of our sample set (Fig. S2). We found similar results when we matched our molecular formulae to those considered universal[21] because they reflect the end-products of degradation (Table S1).

We found evidence that convergence towards universal compounds across soil depth and hillslope was due to progressive reworking of DOM. To assess processing of DOM, we averaged molecular masses in each sample weighted by the relative intensity of formulas[22]. In soils, increasing mass is thought to reflect the transformation of small plant-derived

compounds into larger degradation products by microbes[22]. As expected if DOM shifted from plant- to microbial-derived compounds[22], the weighted molecular mass increased by an estimated mean (95% CI) of 2.5% (<0.1 to 5.1%) from 422 (415 to 430) Da to 433 (426-441) Da from 5 to 60 cm (Table S2). This shift was much stronger from shoulder into stream positions, increasing by 10.1% (4.2 to 16.0%) from 422 (414 to 430) Da to 437 (422 to 452) Da (Table S2). Increases in molecular mass were less consistent when compounds were not weighted by their relative intensities (Table S2).

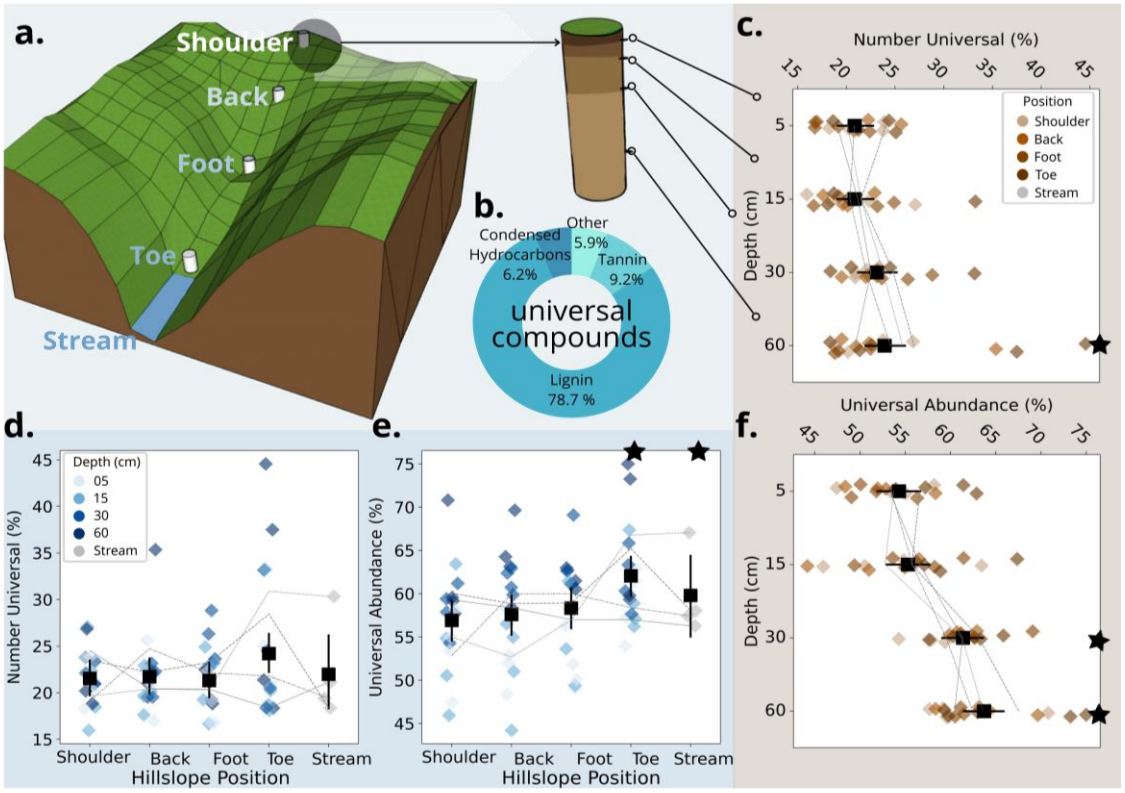

Fig 1. DOM converges to a universal pool across soil depth and hillslope. **a**. Sampling design showing the location of soil lysimeters at shoulder, back, foot, toe hillslope positions and the headwater stream in a study catchment. At each hillslope position, we sampled soil at 5, 15, 30, and 60 cm depth. **b**. Relative contribution of classes to the universal compound pool (n=1216). Lignins and tannins refer to molecular formulae that are phenol derivatives. Estimated marginal means (± 95% CI) for both relative abundance (i.e., sum of normalised

signal intensities) and percentage of compounds comprised by universal compounds along the soil depth (**c.** and **f**., respectively) and hillslope (**d**. and **e**., respectively) gradients. Gradients of colour are increasing soil depth or hillslope position. Lines are mean values for catchment replicates (n=4). Black stars denote hillslope positions and depths that are statistically different from either the shoulder or 5 cm samples, respectively (Table S1).

Convergence towards a universal DOM pool was due to subsets of compounds actively reworked by microbes. Universal compounds classified as lignin-like (representing phenol derivatives) increased from an estimated mean (95% CI) of 40.1% (37.4 to 42.8%) of the summed peak intensities of all molecular formulae to 52.9% (50.2 to 55.6%) from 5 to 60 cm depth and from 43.6 (40.9 to 46.2%) to 48.4% (43.1 to 53.8%) from the shoulder position into the stream (Table S3). These increases were only partly explained by declines in universal molecular formulae from other compound classes (Table S3). Other non-universal compounds must have also become proportionally less abundant, such as because universal lignin-like compounds were produced from non-universal precursors and/or non-universal compounds were preferentially removed.

To explain further the shift towards universal compounds across the two gradients, we examined the relative abundance and mass of non-universal compound classes. The relative abundance of non-universal tannin-like and condensed hydrocarbon-like compounds together declined by an estimated mean (95% CI) of 12.9 (8.9 to 16.9%) and 6.5 (0.2 to 12.9%) along the depth and hillslope gradients, respectively. Both these compound classes tend to reflect plant material rather than microbial products[34,35]. Although other non-universal classes showed small average increases (<4%; Table S4), these could not balance the declines in tannin-like and condensed hydrocarbon-like compounds, as expected if plant material was

generally being degraded. In support of the degradation of specific compound classes causing a shift towards universal DOM, we found that intensity-weighted mass of non-universal compounds changed along the depth and hillslope gradients (Fig. 2). These changes were similar between gradients despite differences in environmental conditions. Carbohydrate-, unsaturated hydrocarbon-, and lipid-like compounds increased in weighted mass by an estimated mean (95% CI) of 16.2% (10.8 to 21.6%), 12.8% (7.8 to 17.7%), and 9.9% (4.7 to 15.0%) from 5 to 60 cm depths, respectively. Carbohydrate- and unsaturated hydrocarbon-like compounds also increased by 14.7% (6.0 to 23.3%) and 12.0% (3.9 to 20.2%), respectively, consistently across hillslope positions from shoulder into the stream (Table S5). Although lipid-like compounds also increased from shoulder into the stream, they did not do so across the other hillslope positions like the carbohydrate- and unsaturated hydrocarbon-like classes (Table S5), potentially reflecting the larger sizes of lipids produced by aquatic primary producers[36]. Together, these results suggested that new, heavier compounds were being created from lighter precursors, as expected if microbes were processing smaller plant-derived compounds into larger microbial products[22].

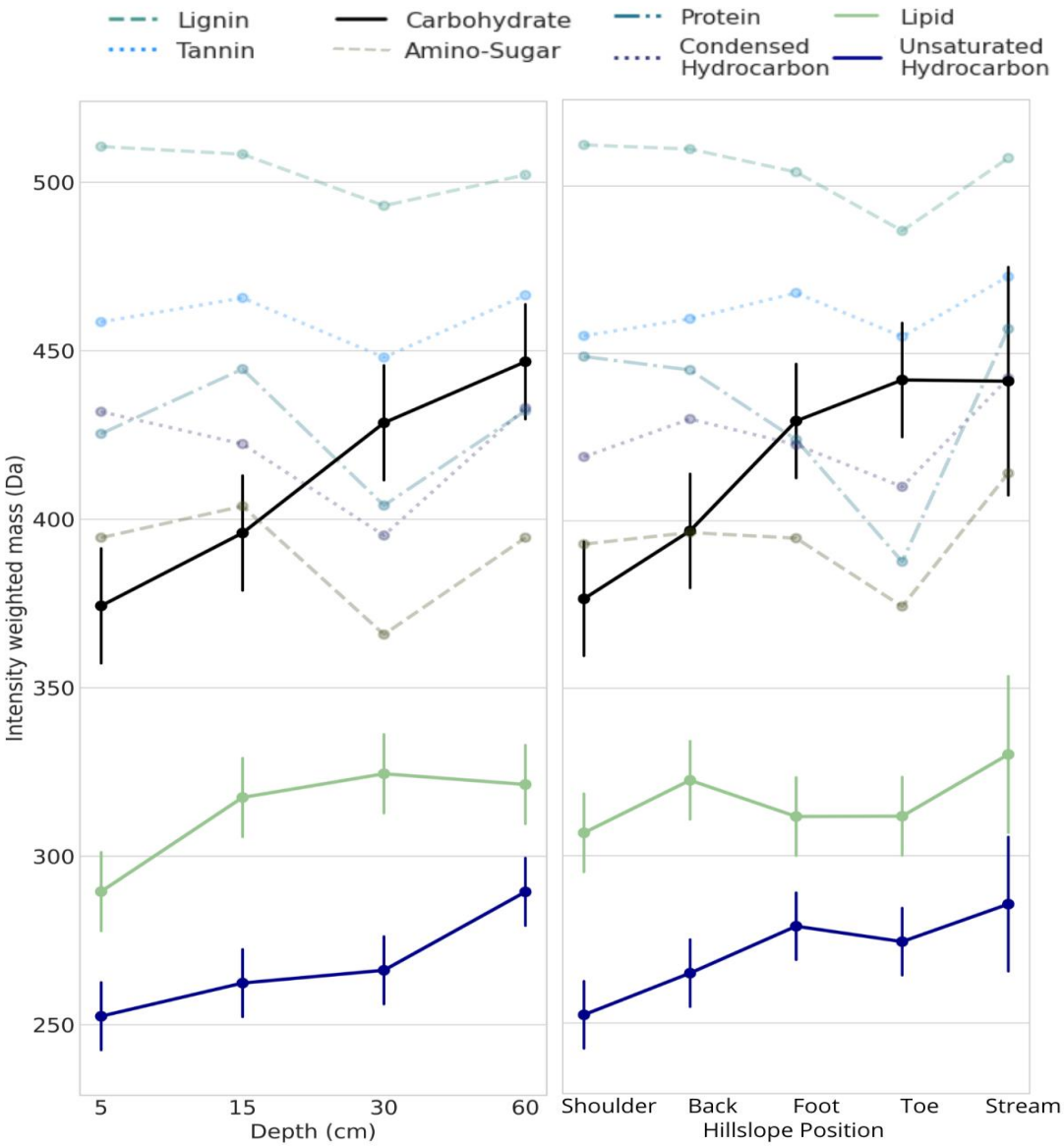

Fig 2. Compound classes shift in mass across soil depth and hillslope position. Mean estimated molecular mass (± 95% CI) of DOM for each compound class when universal compounds were removed. Molecular formulae were grouped into compound classes based on their atomic ratios. Darkened lines are those with statistically significant differences between 5 and 60 cm depths or shoulder and stream positions (Table S5).

Environmental controls over the compound classes were identified by comparing soil depth and hillslope. Absolute carbon concentrations declined with depth and hillslope by an estimated mean (95% CI) 36.8% (24.6 to 52.8%) and 68.0% (47.9 to 94.2%), respectively, reflecting continuous degradation (Fig. 3a). However, the decline could only be attributed to greater microbial productivity along the hillslope (Fig. 3b; Table S6). Consistent with this result, we found a shift from plant- to microbial-derived carbon only along the hillslope using the larger analytical window of size-exclusion chromatography (Table S6). The ratio of humic-like substances to low-molecular-weight carbon chromatographic fractions decreased by an estimated mean of 7.1% (3.4 to 12.6%) towards the streams (Fig. 3c). By contrast, dissolved carbon to total nitrogen (C:N) concentrations in soil porewater decreased with depth, as expected because of microbial processing[37], by an estimated mean of 18.0% (8.1 to 32.3%). There was no change in C:N ratios with hillslope position (Fig. 3d). Alongside the microbial consumption of total carbon (Fig. 3a), nitrogen was likely selectively adsorbed by clays[38] that accumulate at the bottom of hillslopes[39]. In support of this explanation, nitrogen-rich proteins identified by FT-ICR-MS strongly declined from shoulder to toeslope (Fig. 2b). Overall, these results suggest microbial processing explained homogenisation of DOM along the two degradation cascades, although the exact metabolic pathways may have differed.

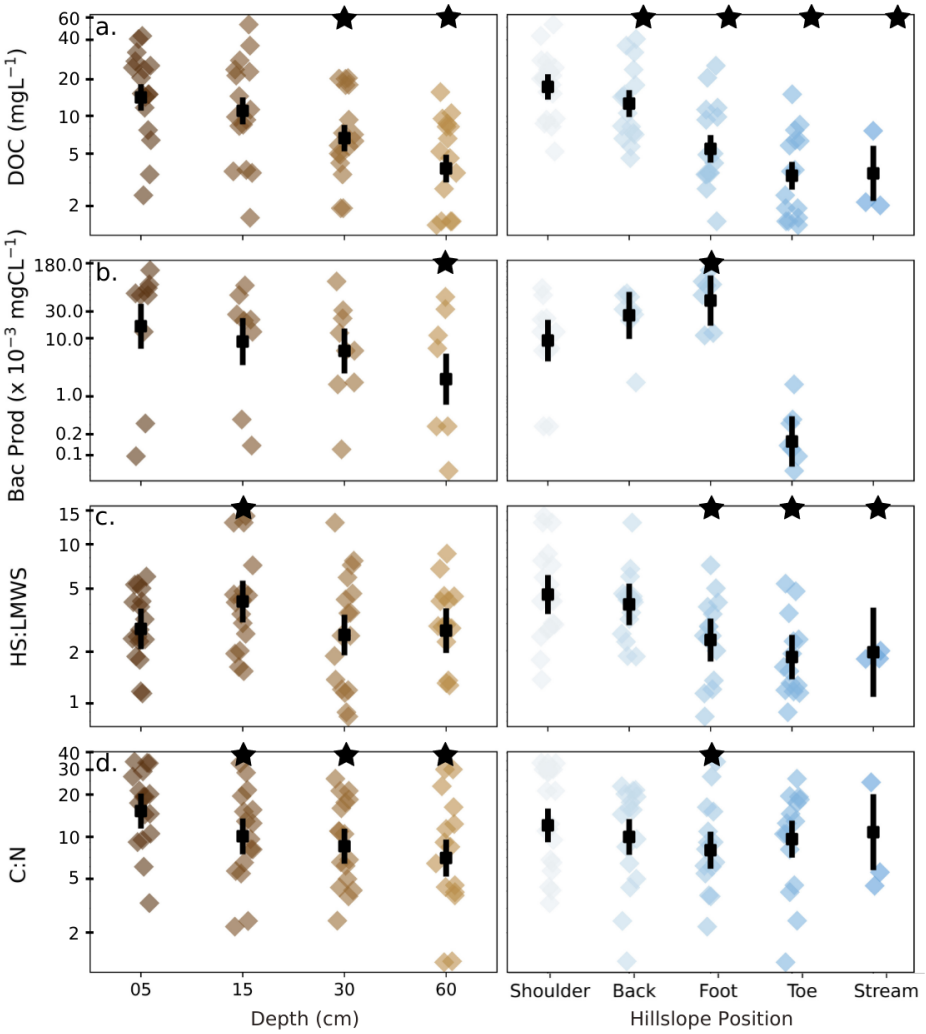

Fig 3. Carbon concentration and quality change along soil depth and hillslope. We measured soil porewater for **a**. dissolved organic carbon (DOC) concentrations. **b**. bacterial protein production (Bac Prod; no stream sample was taken); **c**. fractions of humic substances (HS) to low molecular weight substances (LMWS) using size-exclusion chromatography to expand the analytical window of FT-ICR-MS; and **d.** the ratio of total carbon to total nitrogen (C:N) concentrations. Points are estimated means ± 95% CIs. Black stars denote depths and positions that were statistically different than either the 5 cm or shoulder samples, respectively (Table S6).

Microbial processing explains shifts towards a universal DOM pool

To understand why non-universal compounds in the different classes reflected increased microbial reworking, we partitioned the variation in their relative abundance by soil depth, hillslope position, and environmental variables using redundancy analysis. To prevent overfitting, we reduced the 475 measured environmental variables by removing highly inter-correlated parameters ($|r| > 0.80$) and then further selected predictor variables using an automated permutation-based procedure (see Methods). For carbohydrate- and unsaturated hydrocarbon-like compounds that showed evidence of progressive degradation, that is, increased average molecular mass (Fig. 2), we found that the spatial gradients (depth and position) were more important than for other compound classes (Fig. 4a). Although the percentage of variation explained by these two predictors was small (≤4%), it was comparable on a per-variable-basis to that explained by the 14 to 18 environmental variables included in the final models. Overall, we likely identified many of the most important variables structuring DOM composition as our analysis explained 66 to 91% of the variation in the composition of compound classes.

Most of the environmental variation in DOM composition was due to the functional composition of microbial communities, particularly for compound classes that reflected increased processing along the degradation cascades. We explained differences in the processing of DOM by quantifying the contribution of each environmental variable to the variance partition analysis. This analysis identified 62 environmental variables that were important for explaining DOM composition across the 8 compound classes. For carbohydrate- and unsaturated hydrocarbon-like compounds, which showed increased reworking with both depth and hillslope, variables associated with realised microbial activity together explained more variation in composition than for any other compound class (25 and 27%, respectively; Fig. 4b).

Carbohydrate-active enzymes (CAZymes) primarily used to break down plant-derived carbohydrates were common when we identified the environmental variables shared only by compound classes that changed along the spatial gradients. For carbohydrate- and unsaturated hydrocarbon-like compounds, the activities of four CAZymes were exclusively shared in the lists of the most important environment variables (Fig. 4b). These genes encoded lignocellulolytic enzymes involved in plant cell wall degradation (auxiliary activity 1), glycoside hydrolases involved in degradation of both xylan (glycoside hydrolase 43) and microbial cell walls (glycoside hydrolase 23), and an enzyme involved in the breakdown of polysaccharide carbon-oxygen bonds (polysaccharide lyase 33). We also found evidence that these enzymes were expressed at a community-level. Catabolic use of two carbohydrate substrates (DL-α-glycerol phosphate and β-methyl-D-glucoside) by the microbial communities was also identified exclusively in the lists of the most important variables for carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 4b). For lipid-like compounds that also increased consistently in molecular mass with soil depth, only a family of cellulose-binding enzymes (carbohydrate-binding module 2) was shared with carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 4b). These results provide more direct evidence than previously[22] that microbial processing underlies the progression towards a universal DOM pool along spatial gradients[40].

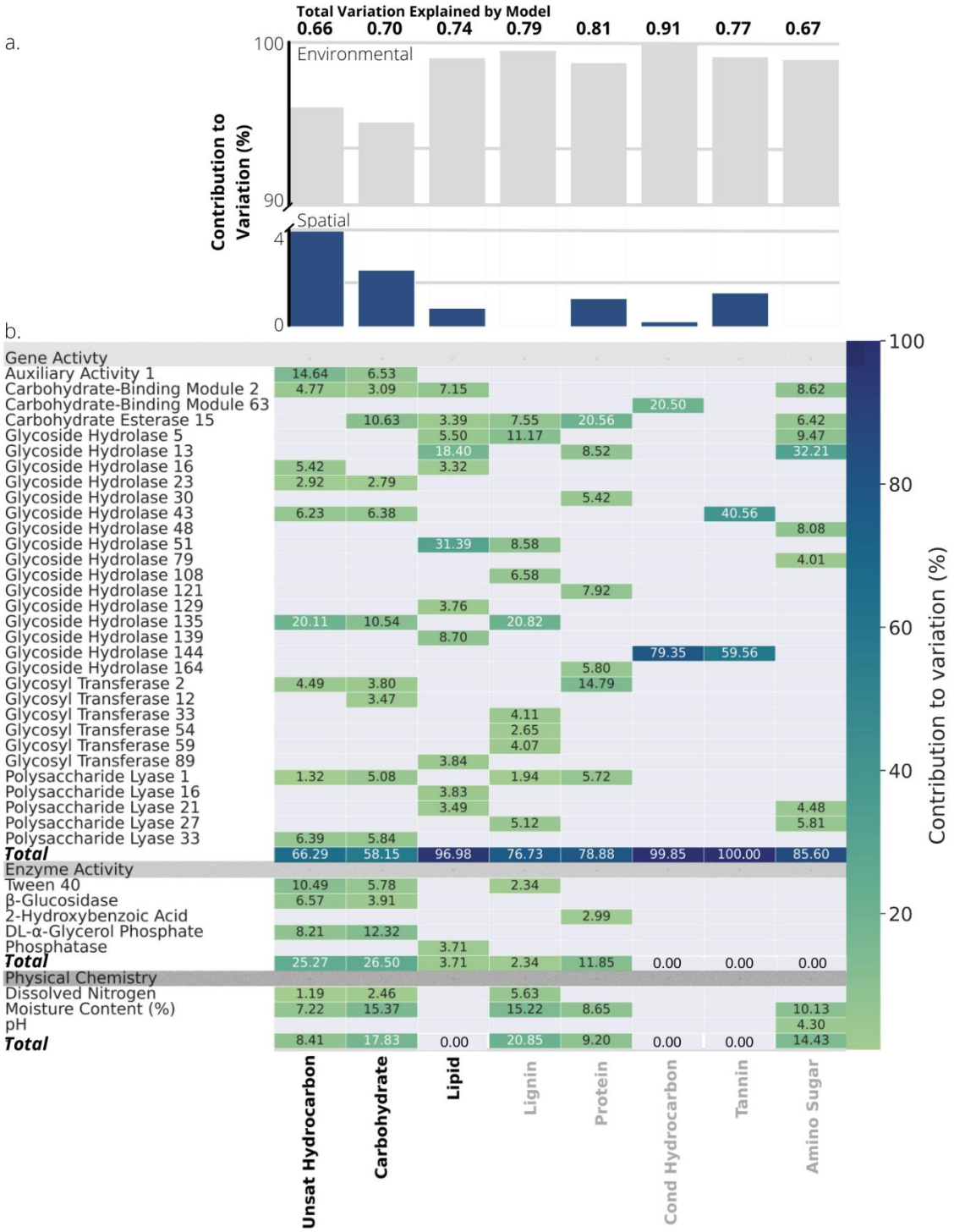

Fig. 4. Spatial gradients and environmental variables explain variation in the molecular composition of different compound classes. **a**. Variation explained by spatial and environmental variables in a variance partition analysis of the relative abundances of non-universal-molecular formula in each compound class. **b**. Relative importance of

environmental predictors in explaining total variance determined by hierarchical partitioning of analysis in **a**. Variables were grouped according to their association with the gene activity of carbohydrate-degrading enzymes, the activity of extracellular enzymes or level of carbohydrate substrate utilisation, and soil physical chemistry. Variation explained by a gene family was displayed if >2.5% for at least one compound class and was summed across all identified subfamilies within a family. Bolded compound classes increased in molecular mass with depth and/or hillslope.

We further found that the soil microbiome universally shifted in activity from processing plant- to microbial-derived OM across the spatial gradients, providing the first direct evidence of its importance in generating a universal DOM pool. We tested if the transcription of genes annotated as CAZymes was differentially expressed. We found that 17 CAZymes statistically differed from 5 to 60 cm depth (Table S7). Of these, only polysaccharide lyase family 1, subfamily 2 (PL1_2) was also associated with classes that increased in molecular mass along the depth gradient (Fig. 4b). PL1_2 is involved in encoding pectate lyase that helps degrade plant-derived OM. As expected if OM inputs shifted from plant- to microbial-derived with increasing soil depth[22], expression of PL1_2 decreased from 5 to 60 cm by a mean (95% CI) of 5-times (2.0 to 11.0). By contrast, 61 CAZymes differed statistically from shoulder to toe positions (Table S8), of which 5 explained about 10%, on average, of the variation in compound classes that showed signs of processing with hillslope (Fig. 4b). The 5 CAZymes included auxiliary activity family 1, which was one of the most important for the composition of carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 4b), as expected if it was involved in oxidising phenols into these classes[41]. The other 4 CAZymes also all increased in expression by a mean of between 2.8 to 3.8-times (1.5 to 6.9) and were associated both with compound classes that

increased (carbohydrate- and unsaturated hydrocarbon-like) and decreased (lignin- and protein-like) in molecular mass from shoulder to toe. These genes included PL1_2 and glycose hydrolases 13 (subfamily 18), 51, and 135, which are associated with degradation of polysaccharides, including celluloses, and fungal biofilm and cell wall components. These results further implicate microbes in transforming plant-derived compounds into higher molecular weight carbohydrates, unsaturated hydrocarbons, and lipids.

Generalising the persistence of DOM across the land-ocean continuum

Our study provides new evidence that DOM converges towards a universal compound pool as microbial reworking removes source-specific components. To determine if this process could apply more generally, we contextualised our observed patterns within the headwater-ocean continuum by synthesising published FT-ICR-MS data. Rather than calculate universal compounds in a global pool of samples, we measured convergence as similarity to a deep-sea reference sample to facilitate inter-study comparison. Although the percentage of compounds shared among all samples in a study reflects convergence towards a universal pool, it is sensitive to differences among studies in molecule number, formula attributions, and sampling intensity. The deep-sea sample is the endpoint of degradation along the land-ocean continuum, so should accumulate the highest proportion of universal compounds[12]. We subsequently found samples expected to be exposed to microbial processing for longer, that is, further along the land-ocean continuum, were most like deep-sea DOM (Fig. 5). The degree of convergence towards a universal pool was consistent with our observations for upland soils, generally, and the position of our different study depths and hillslopes (Fig. 5). These results suggest that a similar process of cumulative exposure to microbial processing may explain changes in DOM along similar spatial and temporal gradients, though the extent to which these trends are linear may vary with time[14], hydrological mixing[26], and rooting

depth[22]. This process would provide the mechanistic basis for popular heuristic models like the river continuum concept[6] and soil chromatograph[42]. More sophisticated ecosystem models could now leverage this pattern in DOM composition to predict variation in DOM degradation[28] and identify potential carbon sinks[43].

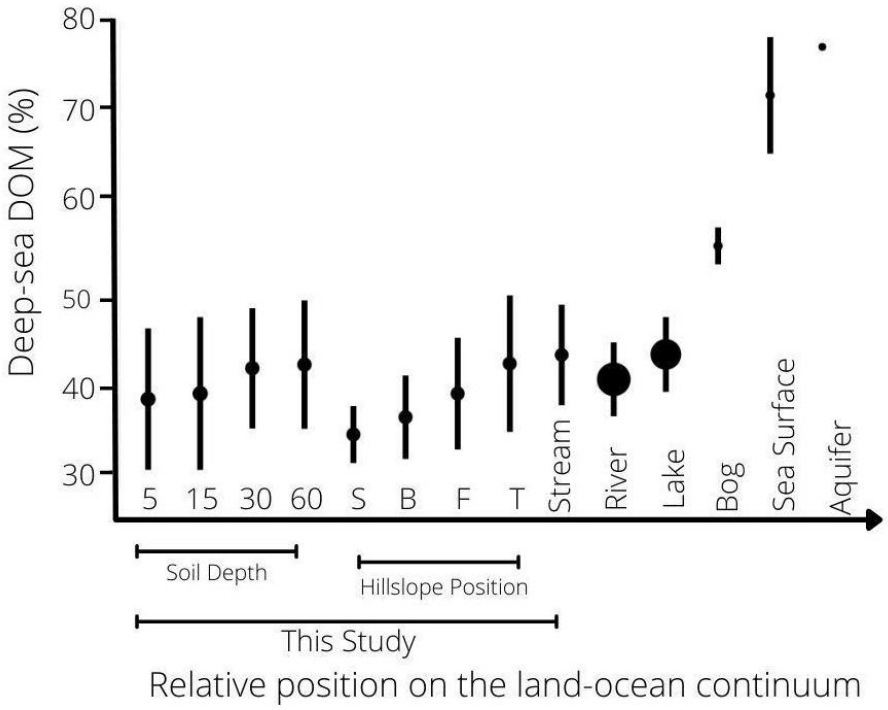

Fig. 5. Soil pore water is an extension of the aquatic continuum. The mean percentage of molecular formulae shared with a deep-sea reference sample was calculated for 433 samples from 5 published FT-ICR-MS studies across a headwater-ocean continuum (Table S9). To account for differences in per-sample molecule number arising from variation in instrument resolution and sample preparation, we used a rarefaction approach employed in microbial diversity studies[44]. We randomly sampled each study 1000 times to a set of 6000 compounds. This threshold was determined based on a minimum number of molecules to adequately sample the pool of observed compounds (Supplementary Methods). Points are

means ± standard deviations of observations. Sizes of points scale with the number of observations at each position along the land-ocean continuum (n = 2 to 116).

Our study advances previous work[22] by demonstrating that microbial processing influences the distribution and thus persistence of organic compounds in soil. Two dominant processes likely contributed towards this result. First, universal compounds, especially those that were lignin-like, may have been selectively retained because of the energetic costs required for their degradation[45]. Second, non-universal compounds were consistently reworked by microbes. Genes encoding enzymes involved in the breakdown of plant cell walls were especially related to mass shifts of microbial-produced compounds, e.g., carbohydrate-like rather than lignin- or tannin-like. Given the ubiquity of plant cell walls in nature, these results suggest widespread microbial metabolic pathways, like those involved in the production of pectate lyases[46], could underpin a general degradation process. This process would persist independently of microbial taxonomic composition and the effects of environmental conditions on certain compound classes, e.g. mineral protection[47]. Even if environmental conditions were modulated along the land-ocean continuum by time or other unmeasured environmental factors, the consistent patterns that we observed across different environments imply a general degradation process. Our study was not designed to test a third hypothesis commonly invoked to explain DOM accumulation in oceans: compounds persist because their concentrations are too low to interact with microbes[24]. Despite homogenisation of chemical traits, DOM has been found to become more structurally diverse along the degradation cascade[48]. Individuals structures may occur at concentrations that limit microbial activity[11], and our methods could be applied in the future to test this idea explicitly. More generally, the processes that we have identified here can help estimate the degradation state of DOM and inform soil-to-stream carbon management.

Methods

Field site and water sampling

Samples were collected from four forested catchments in northwestern Ontario, Canada (46° 58′ N, 84° 22′ W, altitude 375 m; Fig. S1). Within each catchment, hillslopes were partitioned into four positions: shoulder, backslope, footslope and toeslope. Classifications were based on the morphology of surface digital elevation models[49] and a Height Above the Nearest Drainage terrain model[50] that provides a spatial representation of soil-water environments[51]. At all hillslope positions, soil water was sampled in October 2019 after two months of continuous calibration sampling. Samples were collected at 15, 30 and 60 cm depths with tension lysimeters that consisted of 60-mm-long round bottom necked porous cups with an outer diameter of 48 mm and effective pore size of 1.3 µm (model 0653X01-B02M2, Soilmoisture Equipment Corp, USA). The sampling bottles were evacuated to a negative pressure of 50 kPa with a hand pump so suction pressure was ca. 50 mbar above the actual soil water tension. At 5 cm depths, lysimeters could not be securely installed. Therefore, we sampled pore water with micro-tensiometers designed to extract fluids non-destructively using a vacuum[52] through a 0.6 µm ceramic cup (Rhizon CSS samplers, Rhizosphere Research Products, The Netherlands). All the water samplers were installed in triplicate and pooled at each of the depth-by-position combinations to retrieve sufficient volume. Surface water from streams was grab-sampled at the bottom of each hillslope at the channel head. The hillside design of 16 soil samples and 1 stream sample was replicated across the four catchments for a total of 68 samples.

All water was filtered within 8 hours through 0.45 µm glass fibre filters (Whatman GF/F, pre-combusted 400°C, 4 h) and treated differently depending on the eventual analyses.

For fluorescence spectroscopy, samples were aliquoted with no headspace into 20 mL borosilicate scintillation vials and stored at ca. 4°C in the dark. For size-exclusion chromatography, samples were frozen in acid-rinsed and pre-combusted 22 mL borosilicate vials with a PTFE/silicone septa and a polypropylene cap. For DOC concentrations and FT-ICR-MS, samples were acidified with 37% trace metal grade HCl to pH 2 and stored in pre-combusted 40 mL amber borosilicate vials with a PTFE/silicone septa and polypropylene cap at ca. 4°C in the dark.

DOM characterisation and concentration

To characterise DOM, an aliquot of acidified sample (0.1–30 mL, depending on the DOC concentration) was desalted and concentrated via solid-phase extraction (SPE) using a styrene-divinylbenzene copolymer sorbent (100 mg cartridges, Agilent Technologies, USA) using established methods[53]. The sorbent was pre-soaked in HPLC-grade methanol the night prior to extraction. Cartridges were sequentially rinsed with ultrapure water, methanol and ultrapure water acidified with HCl to pH 2. Acidified ultrapure water was stored in the same type of bottles as the procedural blanks. After loading the SPE cartridges with sample, the cartridges were rinsed with acidified ultrapure water and dried. The DOM extracts were eluted with methanol. On each day of extraction, a process blank extract was produced. Redundancy analysis confirmed that the extraction efficiency had no influence on the molecular composition of DOM (explained variation = 0.7%, pseudo-F = 0.6, P = 0.722).

We analysed the DOM extracts on a solariX FT-ICR-MS with a 15 Tesla magnet (Bruker Daltonics, Germany). The system was equipped with an electrospray ionisation source (ESI, Bruker Apollo II) applied in negative ionization mode. We diluted methanol extracts to a final concentration of 2.5 ppm in a 1:1 (v:v) methanol:water solution before

injecting 100 µL into the FT-ICR-MS instrument. For each measurement, we collected 200 scans in duplicate. An in-house mass reference list was used for internal calibration using Data Analysis Software Version 4.0 SP4 (Bruker Daltonics, Germany). The mass error of the calibration was <0.06 ppm for all samples. All measurements were done within 7 days and in random order. In-house reference samples[54] were used to confirm instrumental stability.

Masses ranging from 150 to 1000 m/z were exported, and we assigned molecular formulas using the online platform ICBM-OCEAN[30]. A method detection limit = 2 was applied to all exported masses[55]. All molecules were singly charged ions and we limited formula attributions to $C_{0–100}$, $H_{2–200}$, O_{0-70}, $N_{0–3}$, $S_{0–2}$, $P_{0–2}$ with a tolerance of 0.2 ppm. Only signals detected in both duplicate measurements were retained. We normalised peak intensities of the peaks with an assigned molecular formula to the sum of peak intensities. Intensity-weighted mass-to-charge ratios (m/z_{wm}) for each DOM sample were calculated as the sum of the product of the m/z_i for each compound *i* and their relative intensity I_i divided by the sum of all intensities $m/z_{wm} = \Sigma(m/z_i I_i)/\Sigma I_i$. We also assigned molecular categories to all formula using established criteria[31]. We further defined the structural features[56] in each category to ease semantic confusion among classification schemes (Fig. S3).

A liquid chromatography – organic carbon detection (LC-OCD) system was also used to characterise the DOM pool, including compounds not captured by the analytical window of FT-ICR-MS. With size-exclusion chromatography coupled to a Gräntzel thin-film UV reactor, this procedure differentiated[32]: (i) non-humic "high molecular weight substances" (HMWS) of hydrophilic character; (ii) aromatic humic or "humic-like substances" (HS) with higher aromaticity based on UV absorbance at 254nm; and (iii) "low molecular-weight substances" (LMWS) including acidic and neutral substances. DOC fractions and molecular weights were determined based on compound retention times using the dominating HS peak in the carbon detector chromatograms to identify the position of other fractions. The position

of the HS peak and molecular weight of the HS fraction were calibrated using International Humic Substance Society Suwannee River humic acid and fulvic acid standards[32].

Soil sampling

Microbial activity was measured in soils, as pore sizes of the soil water sampling devices may have excluded soil-bound microbes contributing to the DOM pool. Soils were collected using manual shovelling at each hillslope position in two catchments. Three sterilised 30 cm PVC cores were hammered into an exposed soil wall at each sampling depth (5, 15, 30, and 60 cm). Soil from these cores was homogenised and sieved through a 6 mm mesh to exclude roots and inorganic material using sterile tools. A 10 g subset of soil was cooled on ice for bacterial productivity and respiration experiments performed within 24 hours of sampling. The remaining samples were stored in sterile centrifuge tubes or freezer bags and frozen on-site using dry ice in ethanol. Samples used for sequencing and microbial activity (enzyme assays and substrate use) were stored at -80°C and -20°C, respectively.

Environmental predictors of DOM

We measured 475 environmental variables in soils and their porewaters to partition variation in DOM composition. The variables were associated with microbial metabolic activity and biomass (n=38), microbial diversity (n=6), expression of CAZymes (n=412), and soil water chemical and physical properties (n=19).

We measured 38 variables related to microbial activity and biomass. Bacterial production (BP) was measured using ^{3}H leucine incorporation after Bååth et al[57,58] (Supplementary Methods). Decays per minute measured were converted to BP (mg L^{-1}

day^{-1}) using standard conversion factors[59]. For microbial basal respiration, 1 g of field-moist soil was placed in a 10 mL glass vial and incubated for 24 h in the dark at room temperature (21°C) without manipulating moisture levels. Respired CO_2 (μL min^{-1} g^{-1} dry soil) collected in the headspace of the vial was measured with an infrared gas analyser (QS102, Qubit Systems, Canada). We also measured the potential activity of four hydrolytic enzymes: beta-xylosidase and beta-glucosidase that breakdown xylose and oligosaccharides, respectively, N-acetyl-B-D-glucosaminidase that degrades glycoside and amino sugars, and phosphatase that degrades proteins. All enzyme activities were assayed in 96-well plates under controlled conditions (pH 5, room temperature for 1-hour) using 4-methylumbelliferone-fluorescence tagged substrates and measured with a Synergy H1 Hybrid spectrophotometer/fluorometer (BioTek Instruments, USA) as in existing protocols[60,61]. To complement the enzyme assays, we measured microbial substrate use of 31 different carbon sources using Biolog® EcoPlates™ (Biolog Inc., USA). Cell suspensions were prepared by adding 10 g dry soil equivalent to 95 mL of sterile NaCl solution (0.85%), 6 ceramic beads and 5 g glass microspheres. Samples were mixed on an orbital shaker for 30 minutes at 200 rpm, left to settle for 15 minutes, and a serial dilution on 1 mL supernatant performed to an end concentration of 1:1000 in saline. Inoculated plates were incubated in the dark at 25°C, and absorbance read every 24 hours on the Synergy H1 microplate reader. Blank wells were always subtracted to reduce noise. Average well colour development (ACWD) was calculated as the rate of change over a 24 hr period from day 1 to day 2, as day 2 produced the greatest ACWD indicating the least chance of substrate limitation. Finally, bacterial biomass was measured using flow cytometry. Bacterial cells were separated from soil matrices using buoyant density centrifugation adapted from ref. [62] (Supplementary Methods). Flow cytometry was performed on Accuri™ C6 Plus flow cytometer (BD Biosciences, UK) equipped with a 200 mW solid-state laser emitting light at 488 nm, measuring green

fluorescence at 520 nm (FL1 channel). The FL1 and forward scatter detectors were used to reduce autofluorescence found in environmental samples. We ran samples in triplicate, passing 100 µL per technical replicate through the flow cytometer at a speed of 60 µL min^{-1} to prevent overlap of scatter events. Gates were set by comparing scatter plots produced from stained and unstained samples. Flow cytometer counts were validated with Spherotech 8-peak and 6-peak beads. Bacterial biomass (mg g^{-1} dry soil) was calculated from cell counts, assuming a conversion factor[63] of 58 fg $cell^{-1}$.

Bacterial and fungal taxonomic diversities were assessed using exact sequence variants (ESVs) generated by amplicon sequencing of the 16S rRNA gene and ITS2 region, respectively. DNA was extracted from 250 mg of homogenised soil using the DNeasy PowerSoil Pro kit and the QIAcube® (Qiagen, Germany) automated platform. 16S rRNA and ITS2 libraries were prepared following[64], except that the first set of PCR reactions were set up by mixing 25 µL of HotStarTaq Plus Master Mix, 19 µL RNase-Free Water (Qiagen, homogenised), 0.5 µL of 10 µM primer and 5 µL of gDNA at 5 ng $µL^{-1}$. Indexed and purified amplicons were quantified using the Synergy™ Mx Microplate Reader (BioTek Instruments, USA) before pooling at equimolar concentration. Libraries were sequenced paired-end (2×250 bp) on the Illumina Miseq platform at the Aquatic and Crop Resource Development Research Centre, National Research Council Canada, Saskatoon at an average (±SE) read depth of 23594 (±863) and 34017 (±2380) reads for 16S and ITS, respectively. Sequence data were processed using the MetaWorks pipeline version 1.4.0. Samples were rarefied after ref.[65] to the 15th percentile (7150 and 11103 reads for 16S and ITS, respectively). Any samples with fewer reads were retained if read counts were not 10-times lower than the 15th percentile. Reads were then taxonomically annotated with the RDP classifier v2.13 and UNITE classifier v2.0 for 16S and ITS, respectively. We calculated bacterial and fungal

diversity with the Shannon-Weiner index that accounts for relative abundances of ESVs in addition to their number using the diversity function in the R package vegan[66].

To identify CAZyme genes and quantify their transcripts, we used metagenome and metatranscriptome shotgun sequencing, respectively. Shotgun metagenomic libraries were prepared with the Nextera XT DNA library preparation kit and the Nextera XT Index kit v2 (Illumina, USA) following the manufacturer's instructions using the same input DNA that was used for amplicon sequencing. DNA libraries were purified with Agencourt AMPure XP beads (Beckman Coulter, USA) and fragment size (250–1000 bp) verified on a 2100 Bioanalyzer with a high sensitivity DNA kit (Agilent, USA). Libraries were quantified with the Qubit BR dsDNA assay kit and pooled at equimolar concentrations prior to pair-end sequencing (2×150 bp) at the Centre d'Expertise et de Services Génome Québec on an Illumina Novaseq platform. Metatranscriptomes were obtained by extracting RNA from 2 g of soil using the RNeasy® PowerSoil Total RNA Kit (Qiagen, Germany), except that the phenol/chloroform step was repeated twice. The pellet was suspended in 50 µL RNase/DNase-free water, treated with the RNA Clean & Concentrator-5 with DNase I treatment kit (Zymo Research, USA), and eluted in 15 µL of DNase/RNase-free water. RNA quality was verified with the 2100 Bioanalyzer using the RNA 6000 Nano or Pico assay (Agilent, USA), while RNA concentration was determined with the Qubit RNA high-sensitivity assay kit (Life Technologies, USA). Absence of residual DNA in RNA extracts was further confirmed by PCR amplification of the 16S gene. rRNA was depleted from RNA extracts using the Pan-Prokaryote riboPOOL-kit (siTOOLs Biotech, Germany) with hydrophilic streptavidin magnetic beads (New England Biolabs, USA). rRNA-depleted RNA was then purified with the RNA Clean & Concentrator-5 kit and eluted into 10 µL of DNase/RNase-free water. Libraries were prepared using the NEBNext Ultra™ II RNA Library Prep Kit for Illumina (New England Biolabs, USA) and the NEBNext Multiplex

Oligos for Illumina kit (New England Biolabs, USA) following the manufacturer's protocol for rRNA-depleted RNA. A quality check of the libraries was performed on the 2100 Bioanalyzer with the high sensitivity DNA kit (Life Technologies, USA) prior to pooling and pair-end sequencing (2×125 bp) on an Illumina HiSeq platform at the Aquatic and Crop Resource Development Research Centre, National Research Council Canada, Saskatoon. Metagenomes were screened with Fastp[67] for read adaptor removal and co-assembled per sampling site with metaSpades[68] using Kbase[69] according to default parameters and including the BayesHammer option for read error correction[70]. Gene sequences were identified on the assembled contigs using Prodigal[71] and then annotated as CAZymes using Hidden Markov Models from dbCAN (e-value < 1e-15; coverage > 0.35). Metatranscriptomes were quality-filtered with Fastp according to default parameters[67] and mapped against gene sequences confirmed as CAZymes to obtain their expression profiles using CoverM (https://github.com/wwood/CoverM). Transcript counts were then normalised using the R package DESeq2 to correct for library size and composition and allow for comparison between samples[72].

Finally, major ions, nutrients, and metal concentrations were measured from a subset of the soil pore water at the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario, according to methods outlined in Table S10. Soil moisture content was directly measured by weighing the change in the mass of ca. 10 g of soil before and after drying in an oven for 24 hours at 105°C degrees relative to the original mass.

Statistical analyses

We tested if the molecular composition of DOM varied with soil depth and hillslope position using generalised linear models in R version 4.1.2. The probabilities of detecting universal

molecules, percent relative abundance of universal molecules, and intensity-weighted mass-to-charge ratios (m/z_{wm}) were predicted at each soil depth and hillslope position, while also allowing responses to vary simply because of catchment identity and if the sample was soil or stream. We also allowed depth and position to interact and dropped this term where it was not statistically significant. We used a binomial error structure for models where proportions of counts were predicted and accounted for over-dispersion by including an observation-level random effect. Models were otherwise fitted with a Gaussian error structure and responses that were a proportion of continuous variables were log- or logit-transformed to normalise residuals. Marginal means were predicted at each depth and position by averaging across catchments and either depths or positions using the R package emmeans[73].

To compare the importance of environmental and spatial drivers underlying DOM composition, we used a redundancy analysis (RDA) inferential framework[74]. We partitioned the total unique and shared variation in the relative abundance of each compound class explained by all environmental variables, spatial structure, and the depth and hillslope gradients using RDA ordination estimated with the varpart function from the vegan package[75]. Spatial structure was modelled using principal coordinates of neighbour matrices (PCNM)[76]; see Supplementary Methods. Prior to the variance partition analysis, the 475 environmental predictors were reduced to avoid overfitting. Where two or more variables were correlated with a Pearson $|r| > 0.80$, we removed the variable with the largest mean $|r|$ with all other variables. We further reduced this subset (n=364) for each compound class with stepwise model selection[77] using the ordistep function from the vegan package, dropping variables that were weakly associated ($p > 0.10$) with molecular composition. The importance of individual environmental predictors in each compound class were tested using hierarchical partitioning generalised to multiple predictor matrices implemented with the R package rdacca.hp[78]. All environmental predictors were scaled to zero mean and unit variance.

Data Availability

The compiled chemical dataset used in the analysis is available at: PANGAEA1 and microbial data at: PANGAEA2

DNA and RNA sequence data can be found in the SRA repository under: BioProject PRJNA858277

Acknowledgements

We are thankful to K Waxenberg, W Pallier, and L Mahony (Cambridge) for help with fieldwork. Additional support was provided by K. Klapproth and I Ulber with FT-ICR-MS measurements and C Schmalsch with LC-OCD samples. We also thank staff at the Natural Resources Canada – Canadian Forest Service (NRCan-CFS) for considerable support, including with logistics (K Webster, P Hazlett), development of extraction and assay protocols (V Rouleau, M-J Morency), conducting laboratory extractions and assays (E Smenderovac, D Chartrand), and general field and laboratory assistance (J Schadenberg and many others). This work was funded by a Gates Cambridge Scholarship (OPP1144) awarded to E.C.F, NRCan-CFS Cumulative Effects program awards to E.J.S.E., H2020 ERC Grant sEEIngDOM (804673) to A.J.T, and awards from the Genomic Research and Development Initiative of the Government of Canada to C.M. and E.J.S.E.

References

1. Friedlingstein, P. *et al.* Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
2. *IPCC Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Masson-Delmotte, V.P. et al.) (Cambridge University Press, 2022).
3. Nakhavali, M. *et al.* Leaching of dissolved organic carbon from mineral soils plays a significant role in the terrestrial carbon balance. *Glob. Chang. Biol.* (2020) doi:10.1111/gcb.15460.
4. Catalán, N., Marcé, R., Kothawala, D. N. & Tranvik, L. J. Organic carbon decomposition rates controlled by water retention time across inland waters. *Nat. Geosci.* **9**, 501–504 (2016).
5. Drake, T. W., Raymond, P. A. & Spencer, R. G. M. Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnol. Oceanogr. Lett.* **3**, 132–142 (2018).
6. Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B. & Stackpoole, S. M. The river as a chemostat: Fresh perspectives on dissolved organic matter flowing down the river continuum. *Can. J. Fish. Aquat. Sci.* 150416143426006 (2015).
7. Meyers-Schulte, K. J. & Hedges, J. I. Molecular evidence for a terrestrial component of organic matter dissolved in ocean water. *Nature* **321**, 61–63 (1986).
8. Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E. The river continuum concept. *Can. J. Fish. Aquat. Sci.* **37**, 130–137 (1980).

9. Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804 (2014).
10. Mosher, J. J., Kaplan, L. A., Podgorski, D. C., McKenna, A. M. & Marshall, A. G. Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. *Biogeochemistry* **124**, 371–385 (2015).
11. Seidel, M., Vemulapalli, S. P. B., Mathieu, D. & Dittmar, T. Marine dissolved organic matter shares thousands of molecular formulae yet differs structurally across major water masses. *Environ. Sci. Technol.* **56**, 3758–3769 (2022).
12. Zark, M. & Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 3178 (2018).
13. Dittmar, T. *et al.* Enigmatic persistence of dissolved organic matter in the ocean. *Nature Reviews Earth & Environment* vol. 2 570–583 (2021).
14. Kothawala, D. N., Kellerman, A. M., Catalán, N. & Tranvik, L. J. Organic matter degradation across ecosystem boundaries: the need for a unified conceptualization. *Trends Ecol. Evol.* **36**, 113–122 (2021).
15. Middelburg, J. J. A simple rate model for organic matter decomposition in marine sediments. *Geochim. Cosmochim. Acta* **53**, 1577–1581 (1989).
16. Boudreau, B. P., Arnosti, C., Jørgensen, B. B. & Canfield, D. E. Comment on 'Physical model for the decay and preservation of marine organic carbon'. *Science* vol. 319 1616; author reply 1616 (2008).
17. del Giorgio, P. A., Davis, J., Findlay, S. E. G. & Sinsabaugh, R. L. *Aquatic Ecosystems: interactivity of dissolved organic matter*. (Elsevier, 2003).
18. Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T. & Kattner, G. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochim. Cosmochim. Acta* **69**, 3299–3308 (2005).
19. Hertkorn, N. *et al.* Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010 (2006).
20. Lam, B. *et al.* Major structural components in freshwater dissolved organic matter. *Environmental Science & Technology* vol. 41 8240–8247 (2007).
21. Lechtenfeld, O. J. *et al.* Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **126**, 321–337 (2014).
22. Roth, V.-N. *et al.* Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).
23. Bradford, M. A., Keiser, A. D., Davies, C. A., Mersmann, C. A. & Strickland, M. S. Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. *Biogeochemistry* **113**, 271–281 (2013).
24. Arrieta, J. M. *et al.* Ocean chemistry. Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* **348**, 331–333 (2015).
25. Dungait, J. A. J., Hopkins, D. W., Gregory, A. S. & Whitmore, A. P. Soil organic matter turnover is governed by accessibility not recalcitrance. *Glob. Chang. Biol.* **18**, 1781–1796 (2012).
26. Lynch, L. M. *et al.* River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. *Nat. Commun.* **10**, 459 (2019).
27. Fontaine, S. *et al.* Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* **450**, 277–280 (2007).
28. Zakem, E. J., Cael, B. B. & Levine, N. M. A unified theory for organic matter accumulation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

29. Xenopoulos, M. A., Downing, J. A., Kumar, M. D., Menden-Deuer, S. & Voss, M. Headwaters to oceans: Ecological and biogeochemical contrasts across the aquatic continuum. *Limnol. Oceanogr.* **62**, S3–S14 (2017).
30. Merder, J. *et al.* Improved mass accuracy and isotope confirmation through alignment of ultrahigh-resolution mass spectra of complex natural mixtures. *Anal. Chem.* **92**, 2558–2565 (2020).
31. Kim, S., Kramer, R. W. & Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344 (2003).
32. Huber, S. A., Balz, A., Abert, M. & Pronk, W. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND). *Water Research* **45**, 879–885 (2011).
33. Hawkes, J. A., Hansen, C. T., Goldhammer, T., Bach, W. & Dittmar, T. Molecular alteration of marine dissolved organic matter under experimental hydrothermal conditions. *Geochim. Cosmochim. Acta* **175**, 68–85 (2016).
34. Post-Beittenmiller, D. Biochemistry and molecular biology of wax production in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 405–430 (1996).
35. Haslam, E. *Plant Polyphenols: Vegetable Tannins Revisited*. (Cambridge University Press, 1989).
36. Taipale, S. J. *et al.* Terrestrial carbohydrates support freshwater zooplankton during phytoplankton deficiency. *Sci. Rep.* **6**, 30897 (2016).
37. Rumpel, C. & Kögel-Knabner, I. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
38. Yu, W. H. *et al.* Adsorption of proteins and nucleic acids on clay minerals and their interactions: A review. *Appl. Clay Sci.* **80-81**, 443–452 (2013).
39. Huggett, R. J. *Fundamentals of Geomorphology*. (Routledge, 2007).
40. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
41. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, 41 (2013).
42. Shen, Y., Chapelle, F. H., Strom, E. W. & Benner, R. Origins and bioavailability of dissolved organic matter in groundwater. *Biogeochemistry* **122**, 61–78 (2015).
43. Jaffé, R. *et al.* Global charcoal mobilization from soils via dissolution and riverine transport to the oceans. *Science* **340**, 345–347 (2013).
44. Haegeman, B. *et al.* Robust estimation of microbial diversity in theory and in practice. *ISME J.* **7**, 1092–1101 (2013).
45. Jouanin, L. & Lapierre, C. *Lignins: Biosynthesis, Biodegradation and Bioengineering*. (Academic Press, 2012).
46. Hugouvieux-Cotte-Pattat, N., Condemine, G. & Shevchik, V. E. Bacterial pectate lyases, structural and functional diversity. *Environ. Microbiol. Rep.* **6**, 427–440 (2014).
47. Saidy, A. R., Smernik, R. J., Baldock, J. A., Kaiser, K. & Sanderman, J. The sorption of organic carbon onto differing clay minerals in the presence and absence of hydrous iron oxide. *Geoderma* **209-210**, 15–21 (2013).
48. Zark, M., Christoffers, J. & Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. Mar. Chem. 191, 9–15 (2017).
49. Conacher, A. J. & Dalrymple, J. B. The nine unit landsurface model: an approach to pedogeomorphic research. *Geoderma* **1–2**, 127–144 (1977).
50. Nobre, A. D. *et al.* height above the nearest drainage – a hydrologically relevant new

terrain model. *J. Hydrol.* **404**, 13–29 (2011).
51. Nobre, A. D. *et al.* HAND contour: a new proxy predictor of inundation extent. *Hydrol. Process.* **30**, 320–333 (2016).
52. Seeberg-Elverfeldt, J., Schlüter, M., Feseker, T. & Kölling, M. Rhizon sampling of porewaters near the sediment-water interface of aquatic systems. *Limnol. Oceanogr. Methods* **3**, 361–371 (2005).
53. 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. *Limnology and Oceanography: Methods* **6**, 230–235 (2008).
54. Simon, C., Roth, V.-N., Dittmar, T. & Gleixner, G. Molecular signals of heterogeneous terrestrial environments identified in dissolved organic matter: a comparative analysis of Orbitrap and Ion Cyclotron Resonance mass apectrometers. *Front Earth Sci.* **6**, 138 (2018).
55. Riedel, T. & Dittmar, T. A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **86**, 8376–8382 (2014).
56. Merder, J. *et al.* ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **92**, 6832–6838 (2020).
57. Bååth, E., Pettersson, M. & Söderberg, K. H. Adaptation of a rapid and economical microcentrifugation method to measure thymidine and leucine incorporation by soil bacteria. *Soil Biol. Biochem.* **33**, 1571–1574 (2001).
58. Bååth, E. Thymidine and leucine incorporation in soil bacteria with different cell size. *Microb. Ecol.* **27**, 267–278 (1994).
59. Bade, D., Houser, J. & Scanga, S. *Methods of the Cascading Trophic Interactions* Project. (University of Wisconsin, 1998).
60. Findlay, S. E. G. & Parr, T. B. Chapter 24 - Dissolved Organic Matter. in *Methods in Stream Ecology (Third Edition)* (eds. Lamberti, G. A. & Hauer, F. R.) 21–36 (Academic Press, 2017).
61. 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).
62. Khalili, B., Weihe, C., Kimball, S., Schmidt, K. T. & Martiny, J. B. H. Optimization of a method to quantify soil bacterial abundance by flow cytometry. *mSphere* **4**, (2019).
63. Frossard, A., Hammes, F. & Gessner, M. O. Flow cytometric assessment of bacterial abundance in soils, sediments and sludge. *Front. Microbiol.* **7**, 903 (2016).
64. Rheault, K. *et al.* Plant Genotype influences physicochemical properties of substrate as well as bacterial and fungal assemblages in the rhizosphere of balsam poplar. *Front. Microbiol.* **11**, 575625 (2020).
65. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
66. Oksanen J. et al. vegan: Community ecology package. R package version 2.6-2 https://CRAN.R-project.org/package=vegan (2022).
67. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
68. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* **27**, 824–834 (2017).
69. Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
70. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14 Suppl 1**, S7

(2013).
71. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
72. Love, M., Anders, S. & Huber, W. Differential analysis of count data – the DESeq2 package. *Genome Biol.* **15**, 10–1186 (2014).
73. Lenth, R. emmeans: estimated marginal means, aka least-squares means. R package version 1.4. 7. 2020. (2022).
74. Rao, C. R. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A* **26**, 329–358 (1964).
75. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614–2625 (2006).
76. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* **196**, 483–493 (2006).
77. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008).
78. Lai, J., Zou, Y., Zhang, J. & Peres-Neto, P. R. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. *Methods Ecol. Evol.* **13**, 782–788 (2022).