NERC Open Research Archive sharing our research

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

Article (refereed) - postprint

Roth, Vanessa-Nina; Lange, Markus; Simon, Carsten; Hertkorn, Norbert; Bucher, Sebastian; Goodall, Timothy; Griffiths, Robert I.; Mellado-Vázquez, Perla G.; Mommer, Liesje; Oram, Natalie J.; Weigelt, Alexandra; Dittmar, Thorsten; Gleixner, Gerd. 2019. **Persistence of dissolved organic matter explained by molecular changes during its passage through soil.** *Nature Geoscience*, 12 (9). 755-761. https://doi.org/10.1038/s41561-019-0417-4

© 2019 Springer Nature Switzerland AG

This version available http://nora.nerc.ac.uk/id/eprint/524959/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at http://nora.nerc.ac.uk/policies.html#access

This document is the author's final manuscript version of the journal article following the peer review process. Some differences between this and the publisher's version may remain. You are advised to consult the publisher's version if you wish to cite from this article.

https://link.springer.com/

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos ('the Trademarks') are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

Title:

Persistence of dissolved organic matter explained by molecular changes during its passage through soil

Authors

Vanessa-Nina Roth[1, 2], Markus Lange[1,*], Carsten Simon[1], Norbert Hertkorn[3], Sebastian Bucher[1], Timothy Goodall[4], Robert I. Griffiths[4], Perla G. Mellado-Vázquez[1,5], Liesje Mommer[6], Natalie J. Oram[6,7], Alexandra Weigelt[8,9], Thorsten Dittmar[10,11], Gerd Gleixner[1]

Affiliations

[1]Max Planck Institute for Biogeochemistry, Hans-Knöll-Straße 10, 07745 Jena, Germany

[2]Thüringer Landesamt für Umwelt, Bergbau und Naturschutz, Göschwitzer Str. 41, 07745 Jena, Germany

[3]Helmholtz Zentrum München, German Research Center for Environmental Health, Research Unit Analytical Biogeochemistry (BGC), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

[4]Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, UK

[5]Politechnic University of Sinaloa, Carretera Municipal Libre Mazatlan Higueras km 3, Colonia Genaro Estrada 82199, Mazatlan, Mexico.

[6]Plant Ecology and Nature Conservation Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands

[7]Department of Soil Quality, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands

[8]Institute of Biology, Leipzig University, Deutscher Platz 5e, 04103 Leipzig, Germany

[9]Systematic Botany and Functional Biodiversity, Institute of Biology, University of Leipzig, Johannisallee 21, 04103 Leipzig, Germany

[10]Research Group for Marine Geochemistry (ICBM-MPI Bridging Group), University of Oldenburg, Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl-von-Ossietzky-Str. 9-11, 26111 Oldenburg, Germany

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

*Corresponding author

Main text

Dissolved organic matter impacts fundamental biogeochemical processes in the soil such as nutrient cycling and organic matter storage. The current paradigm is that processing of dissolved organic matter converges to recalcitrant molecules of low molecular weight and high molecular diversity through biotic and abiotic processes. Here we demonstrate that molecular composition and properties of dissolved organic matter continuously change during soil passage and propose that this reflects a continual shift of its sources. Using ultrahigh resolution mass spectrometry and nuclear magnetic resonance spectroscopy, we studied the molecular changes of dissolved organic matter from the soil surface to 60 cm depth in 20 temperate grassland communities on an Eutric Fluvisol. Applying a semi-quantitative approach we observed that plant-derived molecules were first broken down into molecules containing a large proportion of low molecular weight compounds. These low molecular weight compounds decreased in their abundance during soil passage, while larger molecules, depleted in plant-related ligno-cellulosic structures, became more abundant. These findings indicate that the small plant-derived molecules were preferentially consumed by microorganisms and transformed into larger microbial-derived molecules. This suggests that dissolved organic matter is not intrinsic recalcitrant but instead, it is persisting in soil due to simultaneous consumption, transformation and formation.

Dissolved organic matter (DOM) fluxes are an important component of the global carbon cycle[1,2] that enable cycling and distribution of carbon and nutrients[3]. DOM, initially leached from decomposing plant material or exuded directly e.g. as carbohydrates and organic acids from

roots, changes its characteristics during its transport through the soil[3,4]. Thereby, DOM integrates the information of all processes which reach from the vegetation down to the lower limits of groundwater, known as the Critical Zone[5,6]. Thus, organic matter production, its degradation, reprocessing, storage and transport downward all take place in the Critical Zone[7]. These processes have been identified as the cause of rapid changes in terrestrial DOM properties as it passes through the soil profile and finally enters the groundwater. However, a general consensus how these processes affect the molecular DOM properties during its downward transport is lacking. Over the last years a debate about the chemical nature of DOM molecules took place, namely if there are molecules being intrinsically recalcitrant[8,9]. Recalcitrant molecules are assumed to be refractory and not easily decomposed because of their molecular properties. As a consequence of intrinsic recalcitrance, the same molecular structure containing the same atoms remain in the system. However, recent findings suggest that the turnover of soil organic matter with soil depth is largely controlled by its accessibility[10], concentration[11] and also its bioavailability and biodegradability[12,13] but not primarily by an intrinsic recalcitrance of DOM molecules[8,14]. Instead, the concept of persistence[15] suggests that DOM molecules are continuously found in soil, but are constantly recycled, transformed and newly built up. As a consequence of persistence, the same molecular structure holding different atoms are found in the system. Vertical transport of DOM through the soil column is accompanied by a series of sorption and desorption processes in concert with microbial processing[4,12,16]. Aspects of structural changes of DOM with depth are even visible with the naked eye due to preferential loss of light-absorbing structures, i.e. coloured organic matter. Therefore the colour of DOM can change from yellowish-brown to transparent[7,17]. These changes are independent of the decrease of dissolved organic carbon concentration (DOC). At the level of the compound classes, the

hydrophilic fraction and aromatic C and lignin phenols[17,18], i.e. plant-derived organic molecules[19] decrease with soil depth. In turn, altered and reprocessed microbial-derived molecules increase in abundance[4,16]. The stepwise degradation of lignin to water-soluble[20] and then possibly to extended aromatic compounds[21] is thereby suggested as an important contribution to DOM formation[22].

Sorption processes themselves change the DOM composition during soil passage: for instance, lignin degradation products are more efficiently adsorbed than carbohydrates on e.g. clays and other reactive minerals[23]. Mineral-bound organic molecules can be remobilized by percolating surface-reactive plant-derived DOM[4] or acidic root exudates[24]. These changes of DOM are accompanied by microbial processing of organic molecules[4,16]. Though the microbial community is potentially able to decompose the vast majority of DOM molecules, there is a strong influence of DOM molecular properties on the efficiency of its degradation[13]. Thus, the microbial community adapts to and degrades key DOM components depending on their molecular properties at specific depths. This microbial degradation leads to changes in the DOM composition and structure with soil depth[12]. However, microbial processing does not only involve consumption and degradation, but probably also the build-up of new molecules[25,26,27]. Applying non-targeted ultrahigh resolution analytical techniques allows us to capture molecular level information of DOM for a vast variety of molecules. This enhanced level of information describes alteration of molecular functionalities and molecular mass of DOM along gradients and provides access to a more detailed molecular level understanding of the processes than previously available. This insight is needed to link processes that are shown to affect DOM degradation, transformation and (re)synthesis during passage through the Critical Zone and to reveal the relative importance of the proposed mechanisms.

In this study we examine the molecular composition of soil DOM from 20 grassland communities from the surface to 60 cm depth using ultrahigh resolution mass spectrometry (Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, ESI-FT-ICR-MS) and Nuclear Magnetic Resonance Spectroscopy (^{1}H NMR). The composition and properties of DOM identified by these ultrahigh resolution methods are related to known environmental drivers of DOM composition, namely: 1) root standing biomass, a proxy indicative of plant-derived inputs, 2) microbial biomass and genetic diversity as proxies for the community of decomposers and recyclers and 3) the clay content in the soil as a proxy for soil physical constraints. We hypothesize that key drivers of the formation and transformation of DOM are mainly inputs from plants and their microbial conversion products, even superimposing physical processes such as sorption to minerals.

Chemical DOM characteristics at different soil depths

In all soil profiles, DOC concentrations decreased with soil depth (R^2 = 0.41; p-value < 0.001), with the largest decrease observed between 30 and 60 cm depth (Supplementary Table 1). Concentrations of dissolved organic nitrogen (DON) did not change with soil depth (p-value = 0.445), and consequently the DON/DOC ratio increased. In all samples of the 20 grassland communities, 4264 molecular formulae of DOM compounds were identified (see Methods). Along the soil profile, the molecular composition of DOM continuously changed as the relative abundance of low mass DOM molecules (m/z = 150-275) decreased with depth in comparison with those in the middle molecular weight range (m/z = 300-450) (Fig. 1a-d). Thus, the molecular composition of DOM became more dissimilar with larger distance between sampling depths (Supplementary Table 2). However, most of the detected molecular formulae occurred in all depths, but with varying intensities. Therefore, the shift of molecular properties

during downward transport, as revealed by differential mass spectra (see Methods), was maximal between DOM at 10 cm and 60 cm depths (Fig. 1e-g; Table 1). In the differential mass spectra, based on the complete spectra ("all compounds", Table 1), mainly aromatic CHO compounds (aromaticity index AI_{mod}[28,29]) of low molecular mass decreased considerably in their abundance with depth. Compounds that were more abundant at 60 cm were mainly CHO compounds, too, but were characterized by remarkably similar and narrow O/C and H/C ranges in the mid-molecular mass range (Fig. 1h) in which the number of isomers for any given molecular compositions is maximal[30]. These trends were also observed when considering only CHNO and, less specific, only CHOS compounds (Table 1; Fig. 1i,j). The shift in abundances was remarkable consistent around m/z 300 for the complete spectra, CHNO- and CHOS compounds, though the latter ones were less numerous (Table 1).

^{1}H solution NMR spectra confirmed the changes in DOM molecular composition with depth that were demonstrated by FT-ICR-MS. The relative proportions of aromatic and olefinic as well as those of O-alkyl carbon, indicative of carbohydrates and methoxy functional groups, decreased, whereas the amount of saturated groups (aliphatic, sp^3-hybridized carbon) increased with depth (Supplementary Fig. 1). The relative proportions of acetate analogues and carboxyl-rich alicyclic molecules were near constant along the depth profile.

Drivers of the DOM transformation during soil passage

The decline of the low molecular mass compounds towards more prominent mid-weight molecular mass compounds over depth (for all considered subsets: all, CHNO and CHOS compounds) was reflected by a significant increase in the weighted mean of m/z (Fig. 1k; Supplementary Table 3). The shift in molecular mass was accompanied by a considerable decrease in abundance of unsaturated molecules as indicated by the weighted mean of H/C ratio

149 and a decrease of the weighted mean of DOM aromaticity in all compounds and CHNO
150 compounds (Fig. 1l,n). The saturation of CHOS compounds did not change with soil depth,
151 while AI_{mod} increased significantly (Fig. 1l,n). The weighted mean of the O/C ratio of the CHNO
152 compound slightly increased with depth, while the weighted mean of the O/C ratio of all
153 compounds and CHOS compounds did not change with depth (Fig. 1m).

154 The depth effect on the molecular properties and composition of DOM was mostly
155 covered by the decrease of root biomass, soil organic carbon and soil nitrogen along the soil
156 profile, but not by the soil clay content (Supplementary Fig. 2). However, the decrease of root
157 standing biomass, organic carbon and nitrogen along the soil profile were highly inter-correlated
158 (Supplementary Table 4). In order to estimate the relative importance of potential drivers of the
159 DOM transformation, we compared the influences of 1) plant inputs (root standing biomass), 2)
160 soil properties (contents of clay, organic carbon, nitrogen) and 3) the microbial community
161 (bacterial and fungal biomass and their respective genetic diversity). To eliminate the inter-
162 correlation among the variables with depth, this estimation was performed for the topsoil.
163 Variation partitioning showed that the chosen parameters accounted for the molecular DOM
164 variation between 67% and 46% (Supplementary Fig. 3). The soil microbial community was the
165 most important predictor (explained on average 25%) followed by the soil properties (19%) and
166 root biomass (10%, Fig. 2a). In addition, a large part of the variation (13%) was jointly explained
167 by the microbial community and root biomass. Bacteria generally explained more of the
168 variability than fungi (Fig. 2b; Supplementary Fig. 3). However, more than 40% of the DOM
169 variation remains unexplained (Fig. 4, residuals), suggesting that other environmental variables
170 contribute to DOM variability. This unexplained DOM variability might be related to the
171 chemical composition of soil clay, the exudate and root chemistry, influenced by plant

composition and diversity, or the community composition of decomposers and higher trophic levels. Furthermore, a significantly positive correlation of the chemodiversity to the bacterial ($r = 0.46$, $p = 0.048$) but not to the fungal diversity ($r = -0.36$, $p = 0.127$) confirmed the strong relation of the bacterial community to molecular DOM composition. Interestingly, this significant relationship was not related to the diversity of DOM compounds (calculated as Shannon index) but to the dominance structure of the DOM samples (Simpson's index[31]). This positive relation indicates that a more diverse bacterial community results in more evenly distributed DOM composition (Supplementary Fig. 4).

A semi-quantitative approximation of the abundances of individual compounds was applied to evaluate whether the increased weighted mean of m/z with depth is a result of the disappearance of the low molecular mass compounds or of new formation of mid molecular mass compounds. Therefore, FT-ICR-MS signal intensities of individual molecular formulae were scaled to DOC concentrations[32]. This approximation indicated that the increase of the weighted mean of m/z with soil depth was caused by both an absolute decline in abundance of low molecular mass compounds and by an absolute increase in abundance of mid and high molecular mass compounds (Fig. 3a). While the abundance of low molecular mass compounds continuously decreased with soil depth, the mid and high mass compounds became more abundant from 10 to 30 cm and declined below 30 cm. Unexpectedly, the shift in the ratio of low-to-mid-molecular mass compounds (I_{mid}/I_{low}) was strongly related to the degradation state of DOM ($R^2 = 0.71$), described as degradation index I_{DEG} (see Methods[33]) (Fig. 3b). Thus, DOM increased in the state of degradation (higher I_{DEG}) and molecular mass (higher I_{mid}/I_{low}) with soil depth.

Mechanisms underlying the persistence of DOM

The results presented here strongly indicate that there is microbial processing, but not production of recalcitrant products as DOM moves through soils and sediments. The continual shift of molecular composition and properties of DOM during its soil passage furthermore indicates an ongoing transformation of DOM on the molecular level (Fig. 4). The most unexpected change along the soil profile was the shift from low molecular mass to mid molecular mass compounds, which seems to be a general pattern, independent of land use and soil (see Supplementary Fig. 5 for supplementary grassland and forest sites). At 10 cm depth, the most abundant compounds are characterized by a high level of unsaturation, related to polycyclic aromatics ($AI_{mod} \geq 0.67$), including condensed combustion-derived dissolved black carbon (C atoms $n > 15$), and other highly aromatic compounds ($0.66 \geq AI_{mod} \geq 0.50$)[22,28]. The highly aromatic compounds may include polyphenols and polycyclic aromatics with aliphatic side chains[28]. Oxygen poor black carbon e.g. polycyclic aromatic hydrocarbon as precursors of these compounds can be of anthropogenic origin as the field site is close to an area, which was formerly highly industrialized. Microbial alteration of these compounds was shown in a recent study, conducted at the same site[34]. The authors reported considerable amounts of polycyclic aromatic compounds that undergo biodegradation, resulting in low molecular weight oxygenated polycyclic aromatic hydrocarbons. The most abundant compounds at 60 cm depth are characterized as highly unsaturated but not aromatic compounds ($AI_{mod} < 0.50$ and $H/C < 1.5$), which could include residues of microorganisms and degradation products of soil organic matter as well as carboxylic-rich alicyclic molecules (CRAM)[35]. Plant-derived DOM is considered to be more unsaturated and aromatic than microbial-derived substances because of the higher lignin content in plant tissues and other secondary aromatic plant metabolites[36,37]. In consequence, our

218 data suggest that the observed changes in molecular properties are likely driven by a fast
219 microbial turnover[26] of plant-derived DOM during downward transport[4,18], which show a high
220 degree of relative unsaturation and aromaticity (Fig. 4). The results of both FT-ICR-MS and
221 NMR spectra suggest that aromatic substances are rather degradable[20] and could be used as
222 indicators of fresh, plant-derived organic matter, or early products of its decomposition. The
223 even greater decrease of molecular functional groups similar to carbohydrates conforms to
224 expectations[17]. These functional groups might be derived from plant products, such as
225 carbohydrates and cellulose.

226 In our study, the soil clay content had only minor effects on the molecular properties of
227 DOM during its soil passage. This contrasts with many studies reporting a strong impact of soil
228 mineralogy on DOM[3,4,12,15]. This discrepancy might be attributed to the fact that the soil clay
229 content is a good proxy for mineral stabilization on regional or global scales[15], but not on local
230 scales as in our study. The scale-dependency of the mineral-DOM relation is supported when
231 comparing the changes in the DOM mass spectra with soil depth between both sites. While the
232 depth effect found at the Jena site (Fig. 1) was confirmed by DOM analyses from supplementary
233 grassland and forest sites (Supplementary Fig. 5), the mass change from small compounds to
234 mid-molecular mass compounds was more pronounced at the Jena site. This might be due to a
235 finer grain size at the Jena site than at the supplementary site, which has a very sandy soil with
236 probably low sorption capacity (see Methods). This indicates two things: first, the general
237 patterns, i.e. the simultaneous decrease of small compounds (150-275 Da) and increase of mid-
238 size compounds (300-450 Da), is independent of the soil properties. And second, the soil
239 properties affect the strength of this mass shift, indicating the microbial-mineral interaction
240 during the DOM soil passage[4]. The generality of the mass shift, independent of soil properties,

241 are in line with a recent study showing that the biochemical composition of mineral-retained
242 organic matter was similar across four different classes of clay minerals[38]. However, describing
243 the soil based on its texture is a relatively broad parameter to fully assess mineral-organic
244 interactions in soils. Since the strength of mineral-organic interactions is also reflected by the
245 soil organic carbon content[39], we have statistically considered the soil organic carbon content in
246 our analyses, as this probably comprises the mineral-organic interactions. However, we
247 acknowledge that additional laboratory experiments (such as our incubation experiment,
248 Supplementary Fig.6) could also provide detailed mechanistic insights on the mineral-organic
249 interactions, to which DOM is exposed during its passage through the soil profile.

250 Although plant-derived inputs in soil water decreased rapidly with soil depth[16], we found
251 that root standing biomass, which is a proxy indicative of belowground plant-derived carbon
252 inputs, influenced DOM molecular properties along the entire soil profile. This finding supports
253 the observation that microbial processing and reworking is responsible for changes in the
254 molecular properties of DOM with depth[4,16,19]. The shared explained variation in DOM by the
255 microbial community and root biomass, and the prominent role of the microbial community to
256 explain DOM variation indicate a strong microbial imprint and transformation of plant-derived
257 inputs into the DOM pool. Recent findings have shown that the microbial community also adapts
258 to changes in DOM characteristics at different soil depths[12]. This highlights the dual role of
259 microorganism in the carbon cycle in soils, decomposing organic matter while simultaneously
260 recycling and producing new molecules. The shift in the abundance of DOM molecules around
261 300 Da further suggests a selective removal of molecules smaller than 300 Da. Such size-
262 selective processes are well known from the transport of small molecules across membranes of
263 plant roots and soil microorganisms. The outer membranes of Gram-negative bacteria, as well as

mitochondria and plastids, contain water-filled transport channels, so-called porins[40]. Molecules < 600 Da can diffuse through porins with a size-dependent diffusion rate that allows for faster uptake of smaller molecules[41]. Thus, the selective removal of small molecules from fresh DOM might be related to root and microbial uptake through porins[42]. However, because of the increased aliphaticity with soil depth, we propose that microbial uptake is the main driver of the shift in the molecular mass of DOM. The microbial processing of DOM may further explain the decrease in its concentrations with depth, while the concentrations of DON remained at the same level[17]. This suggests a microbial recycling of N or preservation of N-containing molecular structures, while C is partly mineralized. In agreement with the absolute increase of the mid-molecular mass compounds in the upper 30 cm of the soil profile (Fig. 3), this strongly points to the formation of new microbial DOM products with higher average masses during soil passage and not to simple dilution of plant-derived DOM in deeper soil layers because of decreased inputs of plant material. Recently, it has been demonstrated that fresh DOM from primary producers is also characterized by low molecular mass compounds, whereas the so-called "refractory" DOM is clearly shifted to higher masses[43]. The shift to higher molecular masses indicates that the decline of small molecules is related to the progressive degradation of plant-derived DOM and microbial resynthesis of new DOM molecules. These molecules might have high proportions of peptides or proteins thereby explaining the rather constant abundance of DON. We performed a supplementary plant decomposition experiment in which DOM extracted from fresh plant material (Supplementary Fig. 6) was dominated by large, source-specific compounds with a relatively high molecular mass (>500 Da). These large plant-derived compounds disappeared during the first weeks of decomposition. In addition the abundance of low molecular mass compounds and decomposition related compounds increased.

Increasing production of low molecular mass compounds during the early decomposition of plant-derived carbon supports the common opinion that large biopolymers are rapidly degraded during decomposition[44,45]. Our results do not support the paradigm of the build-up of humic polymers[15]. Instead, we suggest that the plant material is extracellularly decomposed to smaller molecules, which then are consumed and in part mineralized or transformed to larger microbial-derived molecules that form a secondary pool of soil organic matter and DOM (Fig. 4). Our observations support the conclusion that low molecular mass compounds contain early decomposition products that might be closely related to the source materials[46]. The emerging higher molecular mass compounds indicate the formation of new molecules from the decomposer community, namely microbial tissues, fragments or products and not random polymerization products. Consequently, the processing of DOM does not primarily lead to recalcitrant molecules, but DOM persists in the Critical Zone as it is constantly decomposed, recycled and newly formed (Fig. 4). The vertical transport of DOM through the soil is assumed to be central to the functioning of soil, e.g. for the formation of soil organic carbon along the soil profile[47]. Many soil processes, important for soil functioning, are driven or mediated by microorganisms. DOM is therefore both a source of nutrients and energy for microorganisms in deeper soils[12] and its molecular characteristics are shaped by the microorganisms above. However, DOM in deeper horizons might be primarily a function of the processes at that depth, i.e. with DOM composition becomes more independent of soil passage, but reflects the complex biogeochemical processes at depth[4]. Therefore the interplay between soil microorganisms and DOM resembles the key to understand the functioning of the Critical Zone, and ultrahigh resolution mass spectrometry provides the means to disentangle this interaction.

References

1. Battin TJ, et al. The boundless carbon cycle. *Nat. Geosci.* **2**(9)**:** 598-600 (2009).

2. Roulet N, Moore TR. Environmental chemistry - Browning the waters. *Nature* **444**(7117)**:** 283-284 (2006).

3. Kalbitz K, Solinger S, Park JH, Michalzik B, Matzner E. Controls on the dynamics of dissolved organic matter in soils: A review. *Soil Sci.* **165**(4)**:** 277-304 (2000).

4. Kaiser K, Kalbitz K. Cycling downwards - dissolved organic matter in soils. *Soil Biol. Biochem.* **52:** 29-32 (2012).

5. Brantley SL, Goldhaber MB, Ragnarsdottir KV. Crossing disciplines and scales to understand the Critical Zone. *Elements* **3**(5)**:** 307-314 (2007).

6. Li L, et al. Expanding the role of reactive transport models in critical zone processes. *Earth Sci. Rev.* **165:** 280-301 (2017).

7. Sanderman J, Baldock JA, Amundson R. Dissolved organic carbon chemistry and dynamics in contrasting forest and grassland soils. *Biogeochemistry* **89**(2)**:** 181-198 (2008).

8. Marschner B, et al. How relevant is recalcitrance for the stabilization of organic matter in soils? *J. Plant Nutr. Soil Sc.* **171**(1)**:** 91-110 (2008).

9. von Luetzow M, et al. Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions - a review. *Eur. J. Soil. Sci.* **57**(4)**:** 426-445 (2006).

10. Dungait JAJ, Hopkins DW, Gregory AS, Whitmore AP. Soil organic matter turnover is governed by accessibility not recalcitrance. *Glob. Change Biol.* **18**(6)**:** 1781-1796 (2012).

11. Don A, Roedenbeck C, Gleixner G. Unexpected control of soil carbon turnover by soil carbon concentration. *Environ. Chem. Lett.* **11**(4)**:** 407-413 (2013).

12. Leinemann T, et al. Multiple exchange processes on mineral surfaces control the transport of dissolved organic matter through soil profiles. *Soil Biol. Biochem.* **118:** 79-90 (2018).

13. Marschner B, Kalbitz K. Controls of bioavailability and biodegradability of dissolved organic matter in soils. *Geoderma* **113**(3-4)**:** 211-235 (2003).

14. Fontaine S, et al. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* **450**(7167)**:** 277-280 (2007).

15. Schmidt MWI, et al. Persistence of soil organic matter as an ecosystem property. *Nature* **478**(7367)**:** 49-56 (2011).

16. Steinbeiss S, Temperton VM, Gleixner G. Mechanisms of short-term soil carbon storage in experimental grasslands. *Soil Biol. Biochem.* **40**(10)**:** 2634-2642 (2008).

17. Kaiser K, Guggenberger G, Haumaier L. Changes in dissolved lignin-derived phenols, neutral sugars, uronic acids, and amino sugars with depth in forested Haplic Arenosols and Rendzic Leptosols. *Biogeochemistry* **70**(1)**:** 135-151 (2004).

18. Gleixner G, Poirier N, Bol R, Balesdent J. Molecular dynamics of organic matter in a cultivated soil. *Org. Geochem.* **33**(3)**:** 357-366 (2002).

19. Gleixner G. Soil organic matter dynamics: a biological perspective derived from the use of compound-specific isotopes studies. *Ecol. Res.* **28**(5)**:** 683-695 (2013).

20. Klotzbücher T, Kalbitz K, Cerli C, Hernes PJ, Kaiser K. Gone or just out of sight? The apparent disappearance of aromatic litter components in soils. *SOIL* **2**(3)**:** 325-335 (2016).

21. Waggoner DC, Chen H, Willoughby AS, Hatcher PG. Formation of black carbon-like and alicyclic aliphatic compounds by hydroxyl radical initiated degradation of lignin. *Org. Geochem.* **82:** 69-76 (2015).

22. DiDonato N, Chen H, Waggoner D, Hatcher PG. Potential origin and formation for molecular components of humic acids in soils. *Geochim. Cosmochim. Acta* **178:** 210-222 (2016).

23. Saidy AR, Smernik RJ, Baldock JA, 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).

24. Keiluweit M, et al. Mineral protection of soil carbon counteracted by root exudates. *Nat. Clim. Change* **5:** 588-595 (2015).

25. Lange M, et al. Plant diversity increases soil microbial activity and soil carbon storage. *Nat. Commun.* **6:** 6707 (2015).

26. Liang C, Schimel JP, Jastrow JD. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2:** 17105 (2017).

27. Miltner A, Bombach P, Schmidt-Bruecken B, Kaestner M. SOM genesis: microbial biomass as a significant source. *Biogeochemistry* **111**(1-3)**:** 41-55 (2012).

28. Koch BP, Dittmar T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid. Commun. Mass. Sp.* **20**(5)**:** 926-932 (2006).

29. Koch BP, Dittmar T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter (vol 20, pg 926, 2006). *Rapid. Commun. Mass. Sp.* **30**(1)**:** 250-250 (2016).

30. Hertkorn N, et al. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Anal. Bioanal. Chem.* **389**(5)**:** 1311-1327 (2007).

31. Magurran AE. *Measuring Biological Diversity*. (Blackwell Publishing, Oxford, 2004).

32. Seidel M, et al. Molecular-level changes of dissolved organic matter along the Amazon River-to-ocean continuum. *Mar. Chem.* **177, Part 2:** 218 - 231 (2015).

33. Flerus R, et al. A molecular perspective on the ageing of marine dissolved organic matter. *Biogeosciences* **9**(6)**:** 1935-1955 (2012).

34. Bandowe BAM, et al. Plant diversity enhances the natural attenuation of polycyclic aromatic compounds (PAHs and oxygenated PAHs) in grassland soils. *Soil Biol. Biochem.* **129:** 60-70 (2019).

35. Hertkorn N, et al. Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**(12)**:** 2990-3010 (2006).

36. Einsiedl F, et al. Rapid biotic molecular transformation of fulvic acids in a karst aquifer. *Geochim. Cosmochim. Acta* **71**(22)**:** 5474-5482 (2007).

37. Fellman JB, D'Amore DV, Hood E, Boone RDe. Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. *Biogeochemistry* **88**(2)**:** 169-184 (2008).

38. Sanderman J, Maddern T, Baldock J. Similar composition but differential stability of mineral retained organic matter across four classes of clay minerals. *Biogeochemistry* **121**(2)**:** 409-424 (2014).

39. Rasmussen C, et al. Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry* **137**(3)**:** 297-306 (2018).

40. Jap B, Walian P. Structure and functional mechanism of porins. *Physiol. Rev.* **76**(4)**:** 1073-1088 (1996).

41. Nikaido H. Transport across the bacterial outer membrane. *J. Bioenerg. Biomembr.* **25**(6)**:** 581-589 (1993).

42. Lehmann J, Kleber M. The contentious nature of soil organic matter. *Nature* **528:** 60-68 (2015).

43. Osterholz H, Niggemann J, Giebel H-A, Simon M, Dittmar T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nat. Commun.* **6:** 7422 (2015).

435 44. Amon RMW, Benner R. Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**(1)**:** 41-51 (1996).

437
438 45. Riedel T, Zak D, Biester H, Dittmar T. Iron traps terrestrially derived dissolved organic matter at redox
439 interfaces. *Proc. Natl Acad. Sci.* USA **110**(25)**:** 10101-10105 (2013).

440
441 46. Benk SA, Li Y, Roth V-N, Gleixner G. Lignin Dimers as Potential Markers for 14C-young Terrestrial
442 Dissolved Organic Matter in the Critical Zone. *Front. Earth Sci.***6:** 168 (2018).

443
444 47. Neff JC, Asner GP. Dissolved organic carbon in terrestrial ecosystems: Synthesis and a model. *Ecosystems*
445 **4**(1)**:** 29-48 (2001).

446

447 **Corresponding Author**
448 Markus Lange, Email: mlange@bgc-jena.mpg.de,

449 **Acknowledgements**
450 We thank U. Gerighausen for sampling and K. Klapproth for technical support with FT-ICR-MS
451 measurements. This work was supported by the "Zwillenberg-Tietz Stiftung" and the Deutsche
452 Forschungsgemeinschaft (DFG) as part of the Critical Zone Observatory "AquaDiva" (CRC
453 1076) and the Jena Experiment (FOR 1451, GL 262/14 and GL 262/19). The International Max
454 Planck Research School for global Biogeochemical Cycles (IMPRS-gBGC) provided the funding
455 for the PhD scholarship of Perla G. Mellado-Vázquez and Carsten Simon. Constructive
456 comments of two anonymous reviewers improved the manuscript considerably.

457 **Author contributions**

458 VNR and GG conceived and designed the study. ML, VNR and GG wrote the main
459 manuscript text. VRN and TD measured and processed MS data, NH measured NMR data. ML
460 and VRN analysed the data. VNR and SB performed the supplementary decomposition
461 experiment and CS analysed these supplementary data and measured and processed the data
462 from the supplementary sites. LM, NJO and AW provided root standing biomass data, PGMV

provided data on microbial biomass, RIG and TG provided data on microbial diversity. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests

Figure captions:

Figure 1: Molecular changes in soil DOM based on FT-ICR mass spectra. a-d, DOM mass spectra at 10, 20, 30 and 60 cm soil depth. The intensity distribution along the mass axis is bimodal. The intensity maximum for the low molecular mass range is highlighted yellow, and blue for the middle molecular mass range. **e-g,** Differential mass spectra of DOM sampled at 10 and 60 cm soil depth (Methods)**,** and **h-j**, the corresponding van Krevelen diagrams. **k-n,** Linear regressions of the soil depth effect on m/z, H/C, O/C, AI_{mod} of DOM. Values are based on all, CHNO or CHNS compounds.

Figure 2: Variation partitioning for potential drivers of DOM transformation. a, Variation explained by soil, roots and microbial community and **b,** by bacteria and fungi only. Diagrams show the averaged explained variations of m/z, H/C, O/C, AI_{mod} and molecular composition (detailed results in Supplementary Fig. 2). Variation partitioning is based on data gathered at topsoil (Methods).

Figure 3: Shift of molecular DOM masses during soil passage. a, Sum of absolute FT-ICR-MS signal intensities for formulae of low, mid and high molecular masses. **b**, Ratio of small-to-

485 mid molecular mass (I_{mid}/I_{low}) of relative intensities are related to an established degradation
486 index of DOM (I_{DEG}[33]). Shaded areas represent 0.95% confidence intervals.
487
488 **Figure 4: Proposed mechanisms for spatial and temporal evolution of DOM molecular**
489 **structures during its soil passage**. Our results indicate that chemical recalcitrance is not the
490 primary mechanism that preserves small DOM molecules from decomposition. Instead we found
491 that the decomposition of DOM molecules increases with depth. Our findings suggest that the
492 persistence of DOM molecules in soil is due to microbial transformation and that DOM
493 consumption is accompanied by formation of new microbial-derived compounds. The
494 consumption, transformation and production by microorganisms initially lead to a preferential
495 degradation of large plant-derived polymers such as lignin, partial mineralization and
496 transformation into a diverse suite of small molecules which are subsequently consumed by the
497 soil microbial community. As indicated by the decreasing aromaticity and unsaturation with soil
498 depth, DOM molecules found in deeper soils are mainly of microbial origin and are either
499 decomposition products or remnants of bacterial necromass.

Tables

Table 1: Molecular properties derived from differential mass spectra shown Fig. 1e-g.

Properties are shown for all, CHNO and CHOS compounds.

Compounds	Depth	m/z	No of formulae	H/C ratio	O/C ratio	AI_{mod}
all	10	165-341	125	0.44-1.00	0.15-0.75	0.45-0.85
	60	335-507	77	1.10-1.40	0.35-0.56	0.14-0.39
CHNO	10	154-370	251	0.42-1.22	0.13-0.71	0.36-1.00
	60	320-494	104	1.05-1.41	0.35-0.56	0.16-0.45
CHOS	10	247-275	5	1.40-1.80	0.50-0.70	0.00-0.23
	60	369-471	31	1.22-1.50	0.39-0.56	0.00

Methods

Field sites and sampling

Soil water samples were collected on a semi-natural grassland with different grassland communities, being part of The Jena Experiment[48]. The site, located in Jena, Germany (50°55' N, 11°35' E, altitude 130 m) close to the Saale River, was converted from grassland to arable field in the early 1960s and ploughed to a depth of about 30 cm. The field site with its current management as grassland was established in 2002. The soil of the field site is classified as Eutric Fluvisol (FAO-Unesco 1997) developed from up to 2 m-thick loamy fluvial sediments, almost free of stones. As expected for a lowland river floodplain setting, sand content correlates with distance from the Saale river (r = 0.95). Close to the river, the topsoil consists of sandy loam, gradually changing into a silt loam with increasing distance to the river. While soil texture varied considerably among the entire field site, the variation in pH (7.1-8.4), soil organic carbon (5-33 g C kg^{-1}) and total soil nitrogen concentrations (1.0-2.7 g N kg^{-1}) was smaller. This study was carried out on a subset (n = 20) of all grassland communities (plots, n = 80), that were fully equipped in 10, 20, 30 and 60 cm depth with glass suction plates (pore size 1 to 1.6 µm, 1 cm thick, 12 cm in diameter; UMS GmbH, Munich, Germany, installed in April 2002)[16]. The investigated plots were aligned in parallel to the river, with soil being dominated by silt (57.3 ± 5.0 SD) while the portions of clay (22.7 ± 2.8) and sand (20.0 ± 7.5) are relatively similar. Soil water was sampled on May 7, 2014 after two weeks of continuous sampling. The sampling bottles were evacuated to a negative pressure of 250 mbar, so that the suction pressure was approximately 50 mbar above the actual soil water tension. Thus, only the soil leachate[49] was cumulatively collected for two weeks.

In addition to the main site in Jena, soil water was analyzed from a site located in the north east of Germany (Linde, Brandenburg 52°32' N 12°39' E, 45 m a.s.l.), with different soil

and land use properties. This site is characterized by a relatively acidic soil pH (forest: 5.78 ± 0.3, grassland 5.05 ± 0.41). Soils have developed on aeolian sands and show mainly features of podzols and cambisols[50]. In November 2014, suction plates, identical to those at the Jena site, were installed on a grassland, and on stands of oak (Quercus robur) and pine (Pinus sylvestris, at depths of 5, 10, 20, 30 and 60 cm). Samples were taken in February 2016. Water sample treatment was conducted as described for Jena samples.

Soil water sample preparation

On subsamples of the soil water samples pH, concentration of DOC (highTOC, sum parameter analyzer, Elementar Analysensysteme GmbH, Hanau, Germany) and dissolved organic nitrogen (DON) were measured. DON was quantified by subtracting the amount of inorganic nitrogen NH_4^+ (ICS-5000, Thermo Fisher Scientific GmbH, Dreieich), NO_2^- and NO_3^- (Dionex DX-500, Thermo Fisher Scientific GmbH, Dreieich) from the amount of total bound nitrogen TNb (TN-100, a1 environsciences Düsseldorf, Germany). The remaining samples were acidified to pH 2 (HCl, p.a.) and stored at 2°C until DOM was concentrated and desalted by solid phase extraction (SPE)[51]. In brief, Agilent Bond Elute PPL SPE cartridges (1 g) that were soaked with methanol overnight were used. Prior to extraction the cartridges were rinsed with ultrapure water, methanol and ultrapure water acidified with HCl to pH 2. Considering the respective DOC concentration determined from the subsamples, the volume of soil water for extraction was adjusted to load 2 mg organic carbon onto the columns. Acidified ultrapure water was stored in the same type of bottles as soil pore water samples to be used as procedural blanks. After loading the SPE cartridges with sample, the cartridges were rinsed with acidified ultrapure water and dried with nitrogen. The DOM extracts were eluted with methanol. At each day of extraction a process blank extract was produced. The average extraction efficiency for soil water DOM was 69% on a carbon basis (s.d. = 6%) for the samples from the main side and 79% ± 7 s.d. for

555 samples from the supplementary site. Redundancy analysis (RDA) confirmed that the extraction
556 efficiency had no influence on the molecular composition of DOM (explained variation = 0.7%,
557 pseudo-F = 0.6, P = 0.722).

558 ***Assessment of soil carbon and nitrogen, roots and the microbial community***

559 Soil samples were taken in April 2012 at each of the plots at the main site to a depth of
560 1 m using a machine driven soil corer (inner diameter of 8.7 cm, Cobra, Eijkelkamp Agrisearch
561 Equipment, Giesbeek, Netherlands). Soil cores were segmented into 5 cm depth sections. Air-
562 dried soil samples were milled and subsamples were analyzed for organic C and total N with a
563 Vario Max and a Vario EL (Elementar Analysensysteme GmbH, Hanau, Germany), respectively.
564 Organic C was determined as the difference between the total C content and the inorganic C
565 content measured after heating the sample to 450°C for 16 h in a muffle furnace[52].

566 Root standing biomass was sampled in 2014, following the protocol of previous years[53].
567 For the root sampling campaign, three soil cores (inner diameter: 4.0 cm, Eijkelkamp Agrisearch
568 Equipment, Giesbeek, The Netherlands) per plot were taken to a depth of 40 cm. Cores were
569 segmented into five layers: 0-5, 5-10, 10-20, 20-30, and 30-40 cm, which were pooled in the
570 field. Samples were stored at 4°C until washing over a 0.2 mm sieve, which took place within 7
571 days. Clean root biomass was dried at 70°C for 72 hours and weighed. Root biomass was
572 calculated as milligrams dry mass per cubic centimeter[53].

573 The microbial community was assessed using the phospholipid fatty acids (PLFA)
574 method to identify bacterial and fungal biomass, and by sequencing for evaluating bacterial and
575 fungal diversity. For both measures identical soil samples were used. In early May 2012, three
576 soil samples per plot were taken with a corer (inner diameter: 4.8 cm, Eijkelkamp Agrisearch
577 Equipment, Giesbeek, The Netherlands) to a depth of 5 cm, pooled. Within 48 hours after
578 sampling the soil was kept at 4°C, sieved to 2 mm, remains of roots were manually removed and

579 the samples were stored at -20°C until further sample processing. PLFA were extracted
580 according to the method of Bligh and Dyer [54] as modified by Kramer and Gleixner [55,56]. As
581 indicator for fungal biomass PLFA 18:2ω6 was used[57,58]. The bacterial biomass was calculated
582 as sum of the PLFA markers 15:0i, 15:0a, c15:0n, 16:0i, c16:1ω7c, c17:0br, 17:0i, 17:0a,
583 17:1ω8, 17:0cy, 18:1ω7, 19:0cy[57].

584 For sequencing analyses of bacterial and fungal communities DNA was extracted from
585 0.3 g of soil using the MoBIO PowerSoil-htp 96 Well DNA Isolation kit (Carlsbad, CA)
586 according to manufacturer's protocols. The dual indexing protocol of Kozich et al. [59], was used
587 for Illumina MiSeq sequencing of the V3-V4 hypervariable regions of the bacterial 16S rRNA
588 gene (primers 341F[60] and 806R[61]); and the ITS2 region for fungi (fITS7f and ITS4r primers
589 sequences[62]). Amplicon concentrations were normalized using SequalPrep Normalization Plate
590 Kit (Thermo Fisher Scientific) prior to sequencing on the Illumina MiSeq using V3 chemistry.
591 Fungal ITS sequences were analyzed using PIPITS[63]) with default parameters as outlined in the
592 citation. Similar approaches was used for analyses of bacterial sequences, using PEAR (sco.h-
593 its.org/exelixis/web/software/pear) for merging forward and reverse reads, quality filtering using
594 FASTX tools (hannonlab.cshl.edu), chimera removal with VSEARCH_UCHIME_REF and
595 clustering to 97% OTUs with VSEARCH_CLUSTER (github.com/torognes/vsearch). Both
596 bacterial and fungal OTU abundance tables were resampled to a minimum of 4000 reads per
597 sample prior to calculating indices of diversity (Shannon index *H'* and Simpson's index on the
598 basis of OTUs) using the *diversity* function in the "vegan" package[64].
599

FT-ICR-MS

For the FT-ICR-MS measurements extract aliquots were diluted to 20 mg L^{-1} organic carbon in ultrapure water/methanol (1:1). The Bruker Solarix FT-ICR-MS (15 Tesla) at the University of Oldenburg (Germany) was used. Samples were continuously injected into the ESI source with a flow rate of 120 µL h^{-1} and an ESI needle voltage of -4 kV in negative ionization mode. 500 single scans with an ion accumulation time of 0.2 s were recorded over a mass range of m/z 150 to 2000 and added to one spectrum. An in-house mass reference list was used for internal calibration.

To only consider significantly measured masses for statistical analyses, several criteria, also described in Malik et al. [65], were applied. First, only m/z values with a signal-to-noise ratio of the maximum of each m/z value ($s/n_{Max,i} > 5$) were considered[66]. To determine $s/n_{Max,i}$ the maximum relative intensity of each m/z value was divided by the noise. Second, only m/z values were kept that were detected in more than one sample. Third, all m/z values with a s/n_{Blanks} ratio ≥ 20 were removed. To determine s/n_{Blanks} the average of signal intensity across all measured blanks was divided by the noise. Fourth, m/z of low intensity ($s/n_{Max} < 20$) were removed if they were detected in less than 20% of the measurements. Mass to charge ratios with assigned molecular formulae passing all criteria described above (4264 different m/z) were isolated from the remaining m/z list and normalized to the sum of intensities. For molecular formula assignment (C, H, O, N, S and P) an in-house algorithm that is based on Koch et al. [67] and Stenson et al. [68] was used. Ions m/z > 660 were not detected in our samples. Only singly charged ions were considered. In consequence, the m/z values represent the molecular mass (Da) of the detected ions. Matlab R2013a (The MathWorks, Inc.) and R[69] were used for data preparation and evaluation.

623 ***NMR***

624 SPE DOM was pooled depth-wise to samples of 3.5 mg DOC, with equal shares of each
625 individual sample (number of samples for 10 to 30 cm: 20, for 60 cm: 14 samples) resulting in
626 one representative sample for each depth. ^{1}H NMR spectra of solid phase (PPL) DOM were
627 obtained with a Bruker Avance III 800 MHz NMR spectrometer (typically 0.5 mg in 300 µg
628 CD_3OD; Bruker 3 mm sealed MATCH tubes). Proton spectra were acquired at 283 K with a
629 5 mm z-gradient ^{1}H / ^{13}C / ^{15}N / ^{31}P QCI cryogenic probe (in CD3OD, Merck, 99.95% ^{2}H).. 1D
630 ^{1}H NMR spectra were recorded using the first increment of the presat-NOESY sequence (nuclear
631 Overhauser effect spectroscopy); solvent suppression with presaturation and spin-lock, 5 s
632 acquisition time, 15 s relaxation delay, typically 64-512 scans, 1 ms mixing time, 1 Hz
633 exponential line broadening.

634

635 ***Data analysis***

636 Weighted means of formula-based characteristics (m/z, AI_{mod} and H/C) were calculated
637 as the sum of the product of the individual information (m/z_i, $AI_{mod,i}$ or H/C_i) and relative
638 intensity I_i divided by the sum of all intensities (e.g. m/z_{WM} = sum (m/z_i * I_i)/$\Sigma(I_i)$). H/C gives
639 information on the saturation and the modified aromaticity index (AI_{mod}) was used to estimate
640 the aromaticity of individual formulae ($AI_{mod} \leq 0.5$: non-aromatic, $0.5 < AI_{mod} < 0.67$: aromatic,
641 $AI_{mod} \geq 0.67$: condensed aromatic)[28]. To relate mass changes to processes, the shift in masses
642 were examined in relation to a degradation index (I_{DEG}[33]). Therefore the normalized intensities of
643 formulae negatively related to degradation (NEG_{Ideg}: $C_{21}H_{26}O_{11}$, $C_{17}H_{20}O_9$, $C_{19}H_{22}O_{10}$,
644 $C_{20}H_{22}O_{10}$, $C_{20}H_{24}O_{11}$) and formulae being positively related to degradation (POS_{Ideg}: $C_{13}H_{18}O_7$,

645 $C_{14}H_{20}O_7$, $C_{15}H_{22}O_7$, $C_{15}H_{22}O_8$, $C_{16}H_{24}O_8$) were summed up. I_{DEG} was calculated based on
646 formula (1)

647 $$I_{DEG} = \sum(\text{intensitiesNEG}_{Ideg}) / \sum(\text{intensities}(\text{NEGI}_{deg} + \text{POS}_{Ideg})) \quad (1)$$

648 Furthermore, the median intensity maximum of the low I_{low} and middle I_{mid} molecular
649 mass range was introduced. For this purpose 1% of the intensities ranked according to intensity
650 for the low (m/z: 150 to 275) and middle (m/z: 300 to 450) molecular mass range was identified.
651 The median of the identified intensities was set as I_{low} and I_{mid}, respectively. To avoid
652 interference from overlap the range from m/z 275 to 300 was neglected. To compare spectra with
653 respect to the intensity of their low and middle molecular weight mass ranges the ratio I_{mid}/ I_{low}
654 was used.

655 To display differences between the mass spectra of samples from 10 cm and 60 cm depth,
656 differential mass spectra were calculated. The differential mass spectrum is the result of
657 subtracting the relative mass peak intensities of the 10 cm measurement from those of the 60 cm
658 measurement. Resulting positive intensities represent formulae that are of higher abundance in
659 10 cm samples whereas negative intensities indicate formulae that are of higher abundance in
660 60 cm ones. To only consider formulae that were of significant higher abundance for each depth,
661 a threshold value (10% of the median of 1% of the highest absolute intensities)[65] was introduced.
662 This threshold value was determined for each differential mass spectrum individually and the
663 average value (1.35, s.d. = 0.06) was applied to all differential mass spectra. The differential
664 mass spectra were calculated for each plot separately. To focus on the common depth dependent
665 trend of all plots, those formulae that occurred in 90% of all differential mass spectra were
666 identified.

667 The analyses for weighted means (m/z, H/C, AI_{mod}) and differential mass spectra were
668 performed for all molecular formulae (4264 different formulae) but also for subsets of formulae
669 that either contained all formulae with at least 1 N (1704 formulae), 1 S (511 formulae). In the
670 main manuscript the datasets based on different formulae were named "all formulae", N-, S- or
671 P-containing formulae.

672 To investigate the general differences in the molecular DOM composition at each soil
673 depth, the Bray-Curtis distances[70] between each depth for each plot was calculated. The averaged
674 Bray-Curtis distances between each depth are given in Supplementary Table 2. Linear mixed-
675 effects models were used considering the repeated measurements on the same plot along the soil
676 profile to statistically test if depth-dependent effects exist independent of the plot identity (*lme*
677 function in the "nlme" package[71]).

678 Variation partitioning analyses[72] were performed based on the comparison of variance
679 explained by linear models including every possible combination of variables being proxies for
680 mineralogy (clay content), plant derived C (root standing biomass) and microbial communities
681 (biomass and genetic diversity). A series of seven models were fitted for each bacterial and
682 fungal community to extract the unique and shared variance for each combination of variables
683 (mineralogy only, plant C only, microbial community only, mineralogy + plant C, mineralogy +
684 microbial community, plant C + microbial community, and all predictors together). Venn
685 diagrams with two factors were displayed using the *compute.Venn* function in Vennerable; Venn
686 diagrams with three factors were displayed using Euler APE for Windows[73].

687 To relate the bacterial and fungal diversity to the chemodiversity of DOM compounds,
688 the Shannon index (*H'*) and Simpson's index (*D*) were calculated based on the measured DOM
689 compounds (molecular formulae) and their relative intensities (ion abundances) using the

690 *diversity* function in the "vegan" package[64]. The relations between microbial diversity and
691 chemodiversity were determined using the Pearson correlation coefficient (*cor.test* function[69] in
692 R[69]).
693
694 *Supplementary incubation experiment, and Orbitrap DOM measurements*
695 *Incubation experiment*
696 Dried and ground plant shoot material (*Bromus erectus, Leucanthemum vulgare agg., Medicago varia*) were inoculated with a mixture of sand (Sigma-Aldrich, baked at 500°C for 4h),
698 based on published experimental setups[74,75]. Plant and soil material were sampled in May 2016,
699 at the Jena site[48]. Mineral soil from < 10 cm depth was floated (200 g and 400 g soil per L,
700 depending on C content; in ultrapure water, for one hour and the supernatant suspension was
701 filtered (0.7 µm, GF/F) to obtain aqueous inoculum. Incubations were done in 50 ml PE Falcon
702 tubes filled with 5 g of pure sand (drainage) and 10.5 g of sand mixed with plant material (plant /
703 sand ratio of 1:20). Contribution of inoculum solution to C stocks was < 0.1%. Samples were
704 kept at 80% water holding capacity over the course of the experiment. Tube lids were closed but
705 allowed air exchange. Incubations were done in triplicate for each time point (0, 1, 2, 3 weeks
706 after inoculation) and blank incubations were carried along. Samples were extracted three times
707 with 30 ml ultrapure water and ultra-centrifuged after each extraction step (3500 min^{-1}). The
708 unified supernatant was vacuum-filtered (700 hPa, 0.7µm GF/F), acidified to pH 2 (HCl, Roth)
709 and extracted by solid phase extraction[51].
710
711 *DOM analyses of supplementary soil water samples and incubation experiment by Orbitrap MS*

Methanol extracts were diluted to 20 mg L^{-1} DOC with ultrapure water and directly infused into an Orbitrap Elite (Thermo Fisher Scientific, Bremen, Germany & Waltham, USA) equipped with a negative-mode electrospray ionization source (ESI). Instrumental settings for the incubation sample set were: Flow rate, 7 µl min^{-1}; spray voltage, 2.65 kV; source fragmentation, 40 eV; capillary temperature, 275°C; S-Lens RF level, 70%; Automatic Gain Control (AGC) setting, 1E6; accumulation time, max. 100 ms; scan range, 150-1500 m/z; transient length, 0.8 ms; nominal resolution, 240.000, scans collected, 300. These settings were slightly changed to measure the soil water sample set from Linde (changed parameters: scan range, 150-1000 m/z; transient length, 1.6 ms; nominal resolution, 480.000, scans collected, 100).

All measurements were done within days and in random order. In-house reference samples[76] were used to check instrumental stability. External calibration was done every day according to the manufacturer protocol. Raw data were averaged in *Xcalibur* (Thermo Fisher Scientific), transformed with *Proteo Wizard*[77] and further processed in mmass[78]. Peak picking was done at 80% peak height and S/N of 5. Further processing followed similar rules as described in section FT-ICR-MS. Additionally, a peak occurrence filter was applied to the data. In the Linde soil water sample dataset, only peaks were included detected in more than 10% of samples. In the incubation dataset duplicate measurements were used and only those signals were kept that had been detected in both measurements[76].

Code availability

The codes used for this study are available on request.

Data availability

The compiled data set used in our analyses is available at https://dx.doi.org/10.17617/3.28 and root standing biomass at https://doi.org/10.1594/PANGAEA.880324. The raw data are available from the corresponding author on request (mlange@bgc-jena.mpg.de).

References

48. Roscher C, et al. The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic Appl. Ecol.* **5**(2)**:** 107-121 (2004).

49. Scheffer F, Schachtschabel P. *Lehrbuch der Bodenkunde*. (Spektrum Akademischer Verlag GmbH, Heidelberg, 2002).

50. Bauriegel A, Kühn D, Schmidt R, Hering J, Hannemann J. Bodenübersichtskarte des Landes Brandenburg im Maßstab 1:300 000. Kleinmachnow/Potsdam.: Landesamt für Geowissenschaften und Rohstoffe Brandenburg; 2001.

51. 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).

52. Steinbeiss S, et al. Plant diversity positively affects short-term soil carbon storage in experimental grasslands. *Glob. Change Biol.* **14**(12)**:** 2937-2949 (2008).

53. Ravenek JM, et al. Long-term study of root biomass in a biodiversity experiment reveals shifts in diversity effects over time. *Oikos* **123**(12)**:** 1528-1536 (2014).

54. Bligh EG, Dyer WJ. A Rapid Method of Total Lipid Extraction and Purification. *Can. J. Biochem. Physiol.* **37**(8)**:** 911-917 (1959).

55. Kramer C, Gleixner G. Variable use of plant- and soil-derived carbon by microorganisms in agricultural soils. *Soil Biol. Biochem.* **38**(11)**:** 3267-3278 (2006).

56. Mellado-Vázquez PG, Lange M, Gleixner G. Soil microbial communities and their carbon assimilation are affected by soil properties and season but not by plants differing in their photosynthetic pathways (C3 vs. C4). *Biogeochemistry* **142**(2)**:** 175-187 (2019).

57. Frostegard A, Baath E. The use of phospholipid fatty acid analysis to estimate bacterial and fungal biomass in soil. *Biol. Fertil. Soils* **22**(1-2)**:** 59-65 (1996).

58. Zelles L. Identification of single cultured micro-orgnisms based on their whole-community fatty acid profiles, using an extended extraction procedure. *Chemosphere* **39**(4)**:** 665-682 (1999).

59. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **79**(17)**:** 5112-5120 (2013).

60. Muyzer G, Dewaal EC, Uitterlinden AG. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**(3)**:** 695-700 (1993).

61. Yu Y, Lee C, Kim J, Hwang S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* **89**(6)**:** 670-679 (2005).

62. Ihrmark K, et al. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **82**(3)**:** 666-677 (2012).

63. Gweon HS, et al. PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol Evol.* **6**(8)**:** 973-980 (2015).

64. Oksanen J, et al. vegan: Community Ecology Package. (2015).

65. Malik AA, et al. Linking molecular size, composition and carbon turnover of extractable soil microbial compounds. *Soil Biol. Biochem.* **100:** 66-73 (2016).

66. Pohlabeln AM, Dittmar T. Novel insights into the molecular structure of non-volatile marine dissolved organic sulfur. *Mar. Chem.* **168:** 86 - 94 (2015).

67. Koch BP, Dittmar T, Witt M, Kattner G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **79**(4)**:** 1758-1763 (2007).

68. Stenson AC, Marshall AG, Cooper WT. Exact masses and chemical formulas of individual Suwannee River fulvic acids from ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectra. *Anal. Chem.* **75**(6)**:** 1275-1284 (2003).

69. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing (2016).

70. Bray JR, Curtis JT. An Ordination Of The Upland Forest Communities Of Southern Wisconsin. *Ecol. Monogr.* **27**(4)**:** 326-349 (1957).

71. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. nlme: Linear and Nonlinear Mixed Effects Models. (2016).

72. Legendre P, Legendre L. *Numerical Ecology*, vol. 20. (Elsevier, Amsterdam-Lausanne-New York, 1998).

73. Micallef L, Rodgers P. eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses. *Plos One* **9**(7)**:** e101717 (2014).

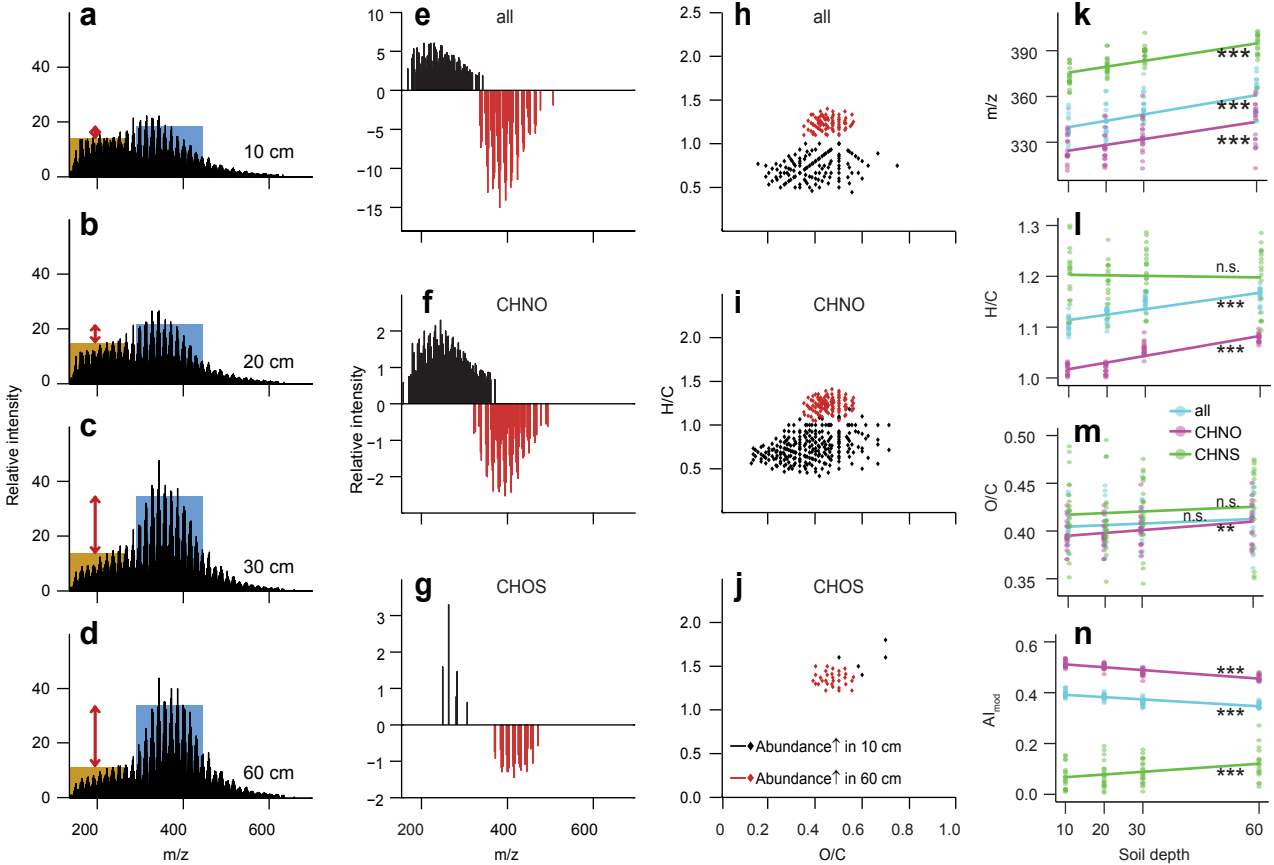
74. Hunt JF, Ohno T. Characterization of fresh and decomposed dissolved organic matter using excitation-emission matrix fluorescence spectroscopy and multiway analysis. *J. Agric. Food Chem.* **55**(6)**:** 2121-2128 (2007).

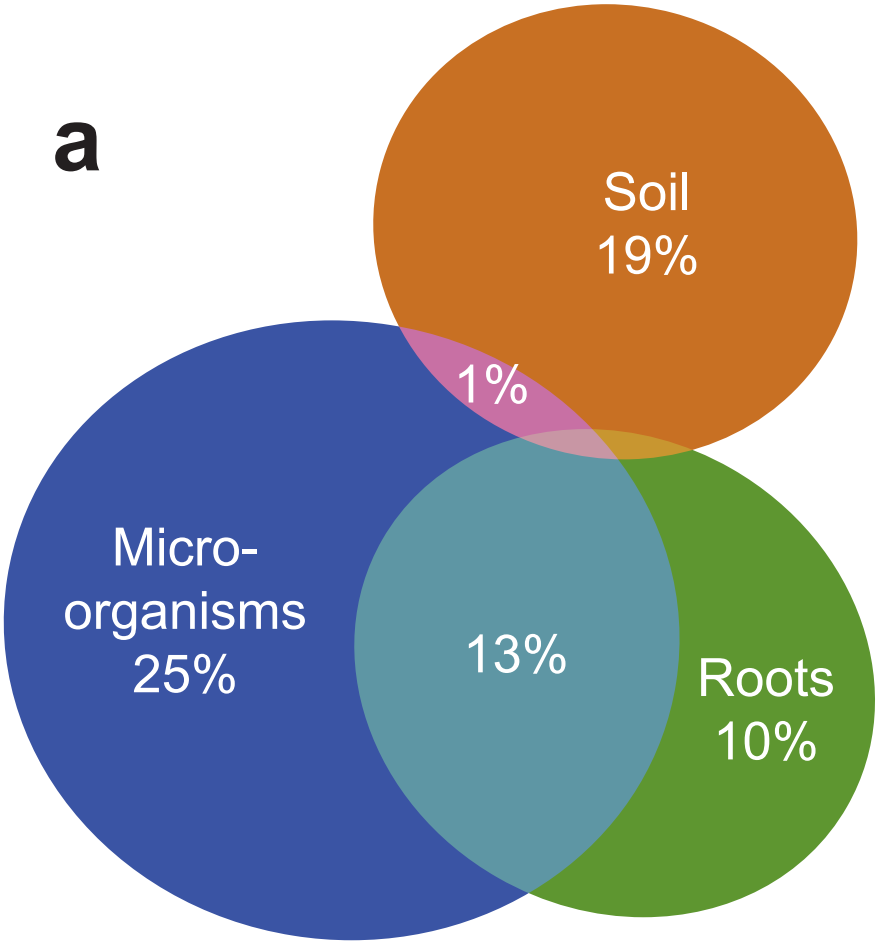
75. Merritt KA, Erich MS. Influence of organic matter decomposition on soluble carbon and its copper-binding capacity. *J. Environ. Qual.* **32**(6)**:** 2122-2131 (2003).

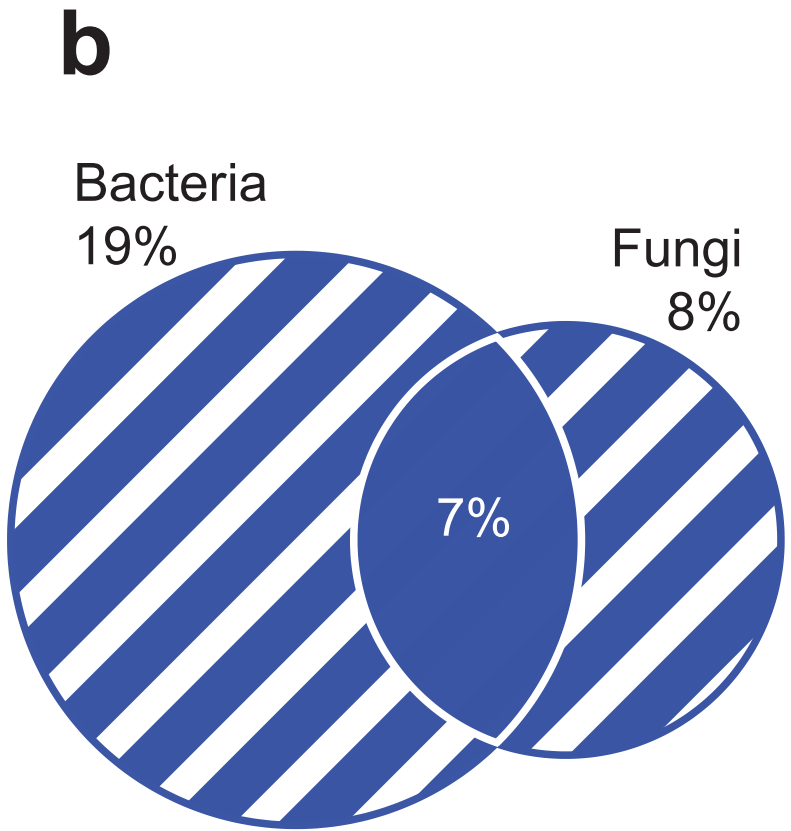
76. Simon C, Roth V-N, Dittmar T, Gleixner G. Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Front. Earth Sci*.**6:** 138 (2018).

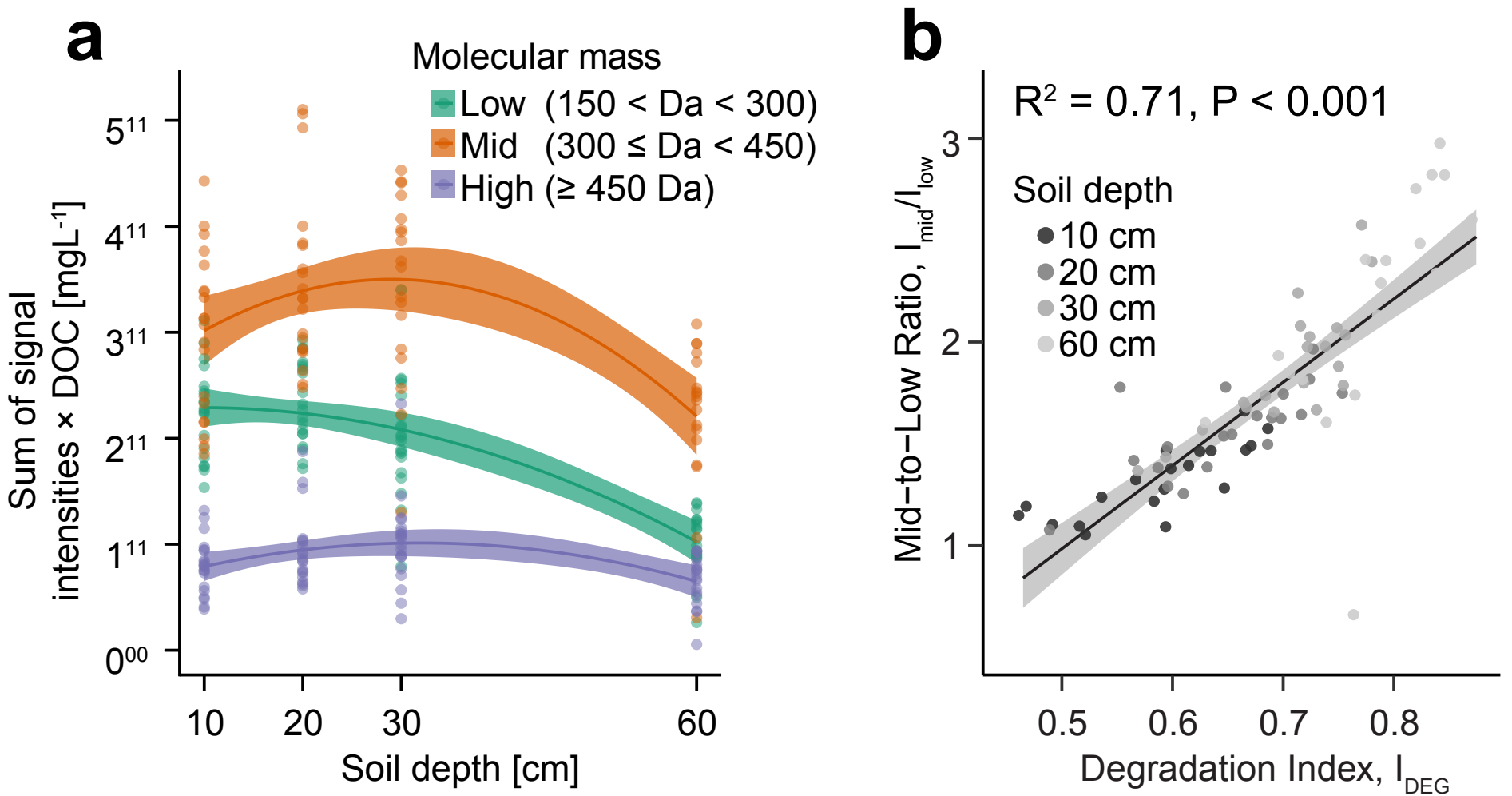
77. Chambers MC, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**(10)**:** 918-920 (2012).

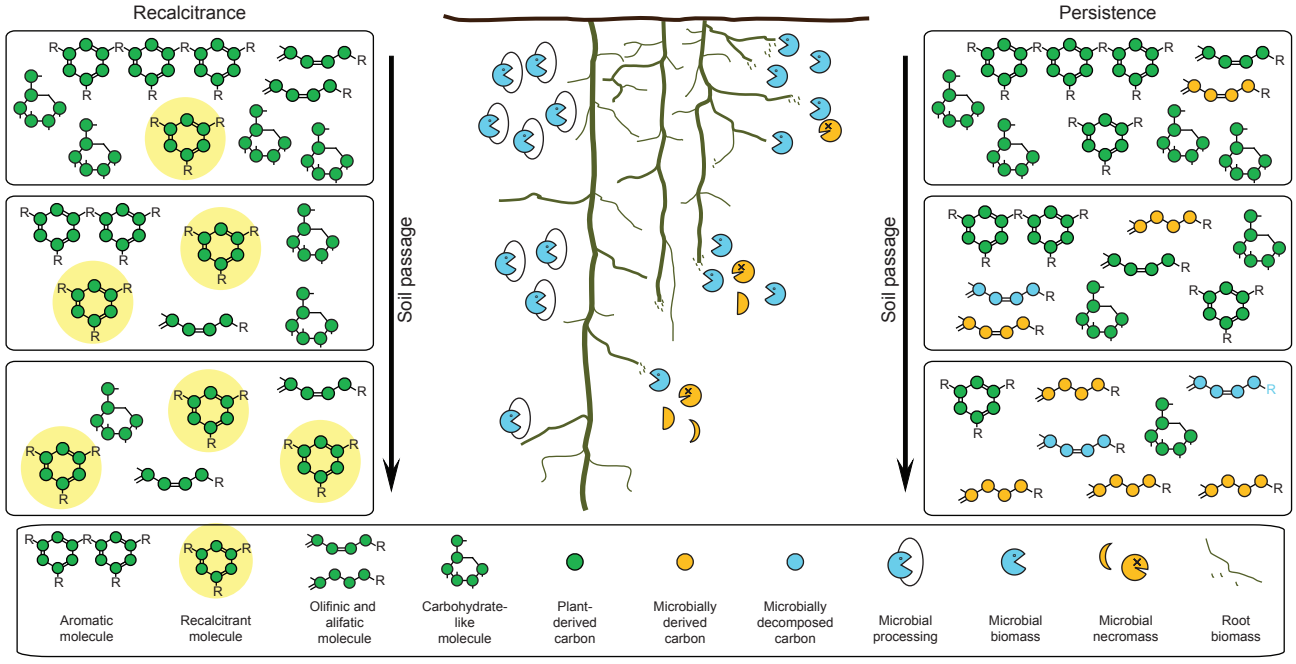
78. Strohalm M, Kavan D, Novak P, Volny M, Havlicek V. mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **82**(11)**:** 4648-4651 (2010).


a
10 cm
b
20 cm
c
30 cm
d
60 cm
Relative intensity
m/z
e
all
f
CHNO
g
CHOS
h
all
i
CHNO
j
CHOS
H/C
O/C
Abundance↑ in 10 cm
Abundance↑ in 60 cm
k
l
m
n
m/z
H/C
O/C
AI mod
n.s.
all
CHNO
CHNS
Soil depth


a
Soil
19%
1%
Micro-
organisms
25%
13%
Roots
10%
Residuals: 43%


b
Bacteria
19%
Fungi
8%
7%
Residuals: 67%


a
Molecular mass
Low (150 < Da < 300)
Mid (300 ≤ Da < 450)
High (≥ 450 Da)
Sum of signal intensities × DOC [mgL-1]
5^11
4^11
3^11
2^11
1^11
0^00
10
20
30
60
Soil depth [cm]
b
R² = 0.71, P < 0.001
Soil depth
10 cm
20 cm
30 cm
60 cm
Mid−to−Low Ratio, I_mid/I_low
3
2
1
0.5
0.6
0.7
0.8
Degradation Index, I_DEG


Recalcitrance
Soil passage
Persistence
Soil passage
Aromatic molecule
Recalcitrant molecule
Olifinic and alifatic molecule
Carbohydrate-like molecule
Plant-derived carbon
Microbially derived carbon
Microbially decomposed carbon
Microbial processing
Microbial biomass
Microbial necromass
Root biomass

Persistence of dissolved organic matter explained by molecular changes during its passage through soil

Figure 1:

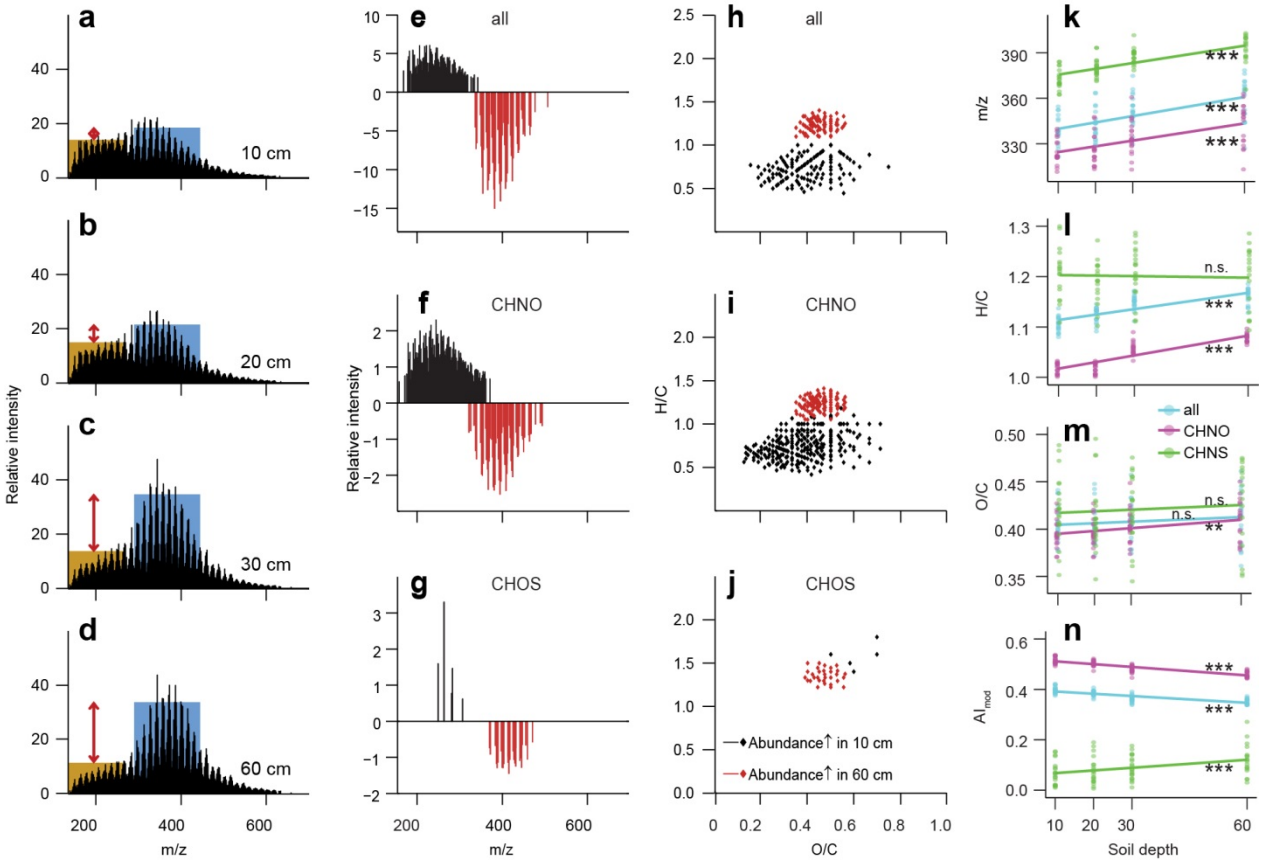


Figure 2:

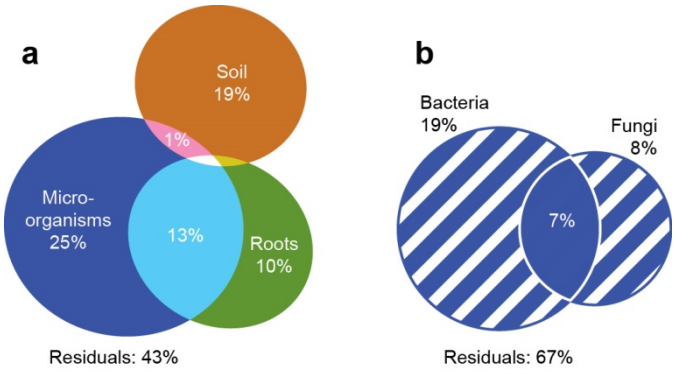

a
Soil
19%
1%
Micro-
organisms
25%
13%
Roots
10%
Residuals: 43%
b
Bacteria
19%
Fungi
8%
7%
Residuals: 67%

Figure 3:

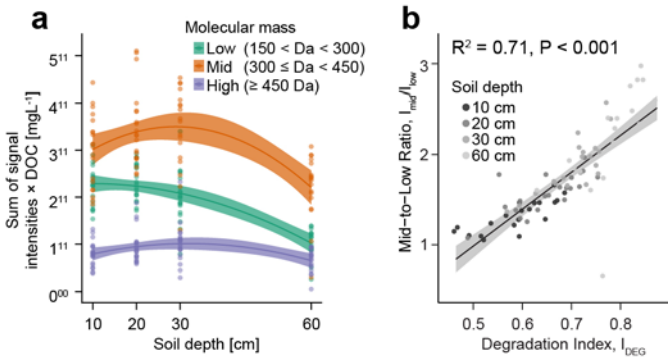
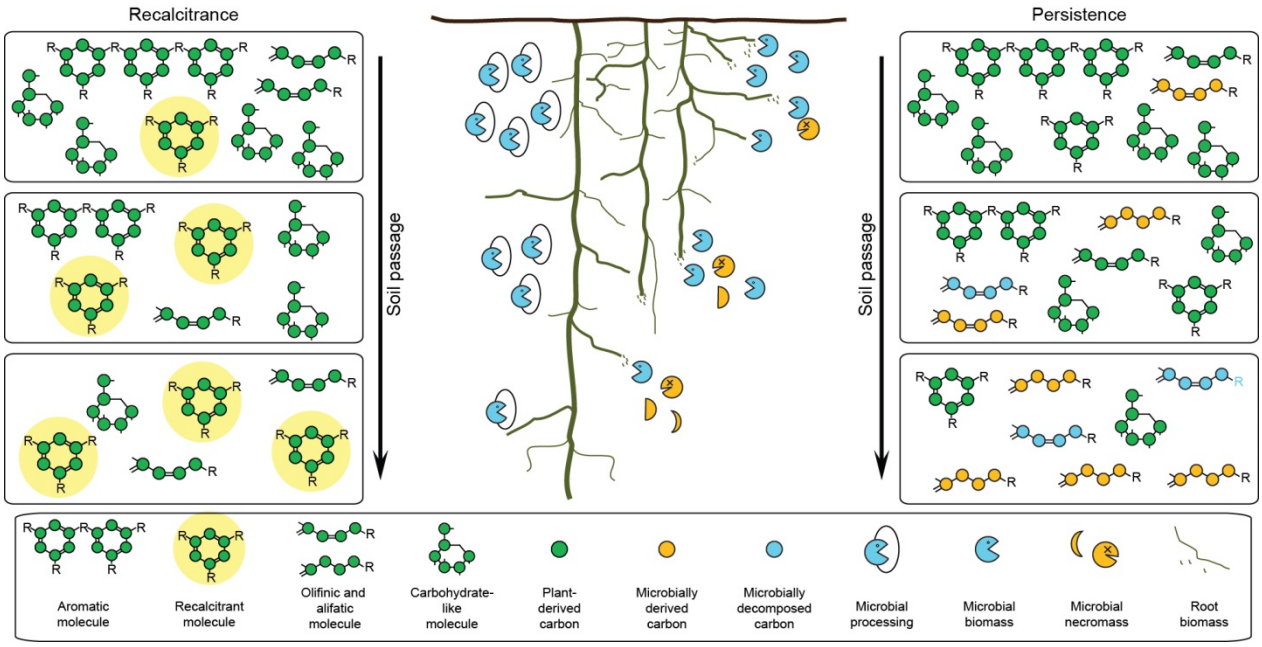

a
Molecular mass
Low (150 < Da < 300)
Mid (300 ≤ Da < 450)
High (≥ 450 Da)
Sum of signal intensities × DOC [mgL⁻¹]
Soil depth [cm]
b
R² = 0.71, P < 0.001
Soil depth
10 cm
20 cm
30 cm
60 cm
Mid–to–Low Ratio, I_mid/I_low
Degradation Index, I_DEG

Figure 4:


Recalcitrance
Soil passage
Persistence
Soil passage
Aromatic molecule
Recalcitrant molecule
Olifinic and alifatic molecule
Carbohydrate-like molecule
Plant-derived carbon
Microbially derived carbon
Microbially decomposed carbon
Microbial processing
Microbial biomass
Microbial necromass
Root biomass