Title

Decomposition decreases molecular diversity and ecosystem similarity of soil organic matter.

Journal

Proceedings of the National Academy of Sciences of USA, 120(25)

Authors

Davenport, Rachelle

Bowen, Benjamin

Lynch, Laurel

et al.

Publication Date

2023-06-20

DOI

10.1073/pnas.2303335120

Peer reviewed

PNAS RESEARCH ARTICLE | AGRICULTURAL SCIENCES

Decomposition decreases molecular diversity and ecosystem similarity of soil organic matter

Rachelle Davenport[a], Benjamin P. Bowen[b,c], Laurel M. Lynch[a], Suzanne M. Kosina[b], Itamar Shabtai[a], Trent R. Northen[b,c], and Johannes Lehmann[a,d,e,f,1]

Edited by Loren Rieseberg, The University of British Columbia, Vancouver, British Columbia, Canada; received February 27, 2023; accepted May 2, 2023

Soil organic matter (SOM) is comprised of a diverse array of reactive carbon molecules, including hydrophilic and hydrophobic compounds, that impact rates of SOM formation and persistence. Despite clear importance to ecosystem science, little is known about broad-scale controls on SOM diversity and variability in soil. Here, we show that microbial decomposition drives significant variability in the molecular richness and diversity of SOM between soil horizons and across a continental-scale gradient in climate and ecosystem type (arid shrubs, coniferous, deciduous, and mixed forests, grasslands, and tundra sedges). The molecular dissimilarity of SOM was strongly influenced by ecosystem type (hydrophilic compounds: 17%, $P < 0.001$; hydrophobic compounds: 10% $P < 0.001$) and soil horizon (hydrophilic compounds: 17%, $P < 0.001$; hydrophobic compounds: 21%, $P < 0.001$), as assessed using metabolomic analysis of hydrophilic and hydrophobic metabolites. While the proportion of shared molecular features was significantly higher in the litter layer than subsoil C horizons across ecosystems (12 times and 4 times higher for hydrophilic and hydrophobic compounds, respectively), the proportion of site-specific molecular features nearly doubled from the litter layer to the subsoil horizon, suggesting greater differentiation of compounds after microbial decomposition within each ecosystem. Together, these results suggest that microbial decomposition of plant litter leads to a decrease in SOM α-molecular diversity, yet an increase in β-molecular diversity across ecosystems. The degree of microbial degradation, determined by the position in the soil profile, exerts a greater control on SOM molecular diversity than environmental factors, such as soil texture, moisture, and ecosystem type.

soil organic matter | molecular diversity | functional diversity

Significance

Diversity indices are widely used in ecology to describe differences in microbial or plant communities, but are rarely applied to investigate variability in decomposing soil organic matter (SOM) composition. We hypothesized that SOM molecular diversity of SOM decreases primarily due to microbial decomposition of plant litter, by transforming dissolved organic matter into molecularly similar microbially derived organic molecules as they move through the soil profile. We found that increased microbial decomposition decreased the total number of molecules and increased the similarity of molecular composition. However, we also found that SOM composition found in deeper soil layers is increasingly dissimilar across different ecosystems, suggesting that microbial decomposition further drives differences across ecosystems.

The molecular diversity of soil organic matter (SOM) has emerged as a potentially critical control on soil organic carbon (SOC) persistence (1). The residence time of SOC influences the relative balance of organic carbon (C) stored in the soil or released to the atmosphere as carbon dioxide (CO_2) and methane (CH_4) (2). The need to fully understand the fundamental controls on SOC decomposition is undeniable. However, the ways and extent to which microbial decomposition may influence organic matter molecular diversity and unique chemical forms of SOM at broad spatial scales have not been studied in any detail.

The fate of SOC is predominantly controlled by soil microbial activity (3), which transforms the composition of plant litter inputs into an array of low-molecular-weight organic compounds (4). Plants contribute organic carbon to the ecosystem via litter deposition on the soil surface (stems, leaves) and belowground (roots and exudates). Root exudates include a suite of primary (e.g., sugars, organic acids, amino acids) (5) and secondary (e.g., alkaloids, phenolics, terpenes) (6) metabolites that directly influence the molecular diversity of SOM. However, a large portion of plant-derived SOM first passes through the microbial "funnel" (7), which fuels microbial metabolism and biomass production. This microbial funnel may directly contribute to molecular diversity by breaking down plant tissue into low-molecular-weight molecules, and indirectly, by releasing a wide variety of exoenzymes, metabolites, extracellular polysaccharides, and lysed microbial cells to the SOM pool. Environmental and chemical pressures may further influence molecular diversity by selecting for microorganisms with the appropriate transporters and metabolic pathways to use a given organic substrate (4). Because plant inputs decrease with depth, SOM is increasingly decomposed as it passes through the soil profile (8–10). Thus, SOM composition directly effects its consumption by microorganisms and vice versa. While some evidence suggests that microbial decomposition increases dissolved organic matter molecular diversity in aquatic ecosystems (11–13), little direct evidence exists for soils (14).

SOC cycles are largely controlled by temperature and moisture, both of which affect microbial and plant community composition (15). The dominant vegetation type in an ecosystem strongly influences the chemical composition (e.g., C:N ratio) of organic matter

Author contributions: R.D. and J.L. designed research; R.D. performed research; R.D., L.M.L., and I.S. contributed new reagents/analytic tools; R.D., B.P.B., S.M.K., T.R.N., and J.L. analyzed data; and R.D., L.M.L., and J.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: CL273@cornell.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2303335120/-/DCSupplemental.

Published June 12, 2023.

inputs to the soil system. Litter with higher N contents generally exhibits greater mineralization rates compared to those with lower N content (16). Despite more rapid initial decomposition, plant litter with higher N content is theorized to increase C storage in mineral–organic associations by increasing microbial carbon use efficiencies (16–19). Plant community composition is also predicted to influence how C is cycled belowground as root architecture and distribution within the soil profile varies among species, influences soil aggregation, and alters microbial community structure and function in rhizosphere versus bulk soils (20–22). What remains largely unexplored is whether the molecular diversity of SOM is more strongly driven by climate conditions, plant litter type, litter decomposition, or soil properties. Specifically, it is not known whether type of plant input or microbial decomposition processes exert a stronger control on molecular diversity.

In environmental sciences, the concept of diversity is primarily used to describe organisms. The advent of high-resolution mass spectrometry has dramatically improved our ability to apply the same diversity tools to assess the molecular diversity of natural organic matter. The mathematical equations used to quantify and describe molecular diversity are largely guided by theory developed from plant, animal, or microbial community ecology (23–25). Various mathematical functions span a wide range of uses and frequently incorporate the abundance and evenness of counted individuals (e.g., species or molecules) in a given population (organismal or SOM pool) (Table 1) (23, 25, 26). However, it is not clear that ecological diversity indices are useful descriptors of organic molecules in soil. With the exception of some studies in aquatic ecosystems (9, 12, 13, 27–29), little work has been done to examine the utility of various mathematical approaches to capture the diversity of molecules in soil (30, 31).

Therefore, our main objective was to identify how climatic variables, plant litter and ecosystem type, and soil properties influence the molecular diversity of SOM within depth profiles that capture gradients in SOM decomposition. We compared six distinct ecosystems with different dominant vegetation classes, soil type, and climate characteristics across the United States, including arid shrublands, coniferous forests, deciduous forests, grasslands, mixed forests, and tundra sedges (n = 3 for all ecosystems). We hypothesized that 1) microbial transformations exert a greater control on SOM diversity than plant litter type, and 2) microbial decomposition of SOM transforms its composition toward a pool of molecularly similar compounds, thus decreasing the diversity of compounds therein.

To identify differences in molecular diversity across ecosystems in the soil profile, we used tandem liquid chromatography mass spectrometry (LC–MS/MS) to analyze i) hydrophobic (nonpolar) metabolites, which were retained and separated using reverse-phase C18 chromatography and ii) hydrophilic (polar) metabolites using hydrophilic interaction liquid chromatography (HILIC). Hydrophobic and hydrophilic molecules represent those that are either immobile or mobile, respectively, via transport by water throughout the soil profile. LC–MS/MS data were used to define unique metabolite features based on exact mass and retention time. Diversity indices were calculated from these features by converting peak heights to relative abundances. We also used putatively identified molecules for each metabolite feature based on MS/MS similarity [Global Natural Products Social Molecular Networking (GNPS) top-hit] as unique chemical "species" for functional diversity analysis. With this information, we tested the suitability of three ecological indices describing either the molecular α-diversity or β-diversity of SOM: 1) molecular richness; 2) abundance-based molecular diversity, which accounts for the richness and evenness of molecules (Hill Numbers) (23, 25); and 3) trait-based functional molecular diversity (Rao's quadratic equation) (34, 36) by determining the nominal oxidation state of carbon (NOSC) and molecular weight for each of the putatively identified metabolites. NOSC and molecular weight indices have been used in an aquatic study to identify functional molecular diversity (27) as both metrics describe the ability of microorganisms to interact with or use SOM. Using these diversity indices, we identified how the molecular α-diversity, or number of distinct molecules, and their trait-based molecular dissimilarity varied with soil depth within individual ecosystems. We also tested how the molecular β-diversity varied across ecosystems using a functional diversity index and the Bray–Curtis Dissimilarity Index. We examined common and unique molecular features across the six ecosystems and their horizons by summing positively identified features. We further assessed organic matter composition using complementary techniques: attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy of SOM and1s C near-edge X-ray absorbance (NEXAFS) spectroscopy of bulk soil and litter chemistries.

Results

Molecular Diversity in Relation to Microbial Decomposition. SOM found deeper in the soil profile is typically older and more microbially transformed than SOM in the topsoil (8, 37–40). Here, we show that the number of common hydrophilic compounds (i.e., annotated molecular features identified in all ecosystems) was four times higher in the litter layer (471 compounds, 31%) than that of the subsoil C-horizon (118 compounds, 13%) (Fig. 1). Within the litter

Table 1. Ecological and molecular usage of diversity indices

Diversity index	Equation	Ecological and molecular context
α-diversity index		
Molecular richness (D_R)	$D_R = S$	Counts known species or molecules (32, 33). More easily identified for organisms, compared to molecular features identified in LC–MS/MS spectra.
Molecular diversity using Hill Numbers($D_{HN,\ q=2}$)	${}^{q=2}D_H(p) = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)}$	Estimates effective number of species or molecules (23, 25). Molecular richness and abundance are considered; however, more weight is given to rare molecules.
β-diversity index		
Functional molecular diversity using Rao's quadratic entropy (FD_{Rao})	$FD_{Rao} = \sum_{i,j=1}^{S} d_{ij}p_i p_j$	Quantifies trait-based variability in a community of species or pool of molecules (27, 34, 35). Traits used for functional diversity of organisms are often specific genes. For SOM molecules, a specific property must be chosen, such as the oxidation state of carbon or molecular weight, which was used in the work.

S is the total number of molecules; p_i is the relative abundance of the *i*-th molecule; p_j is the relative abundance of the *j*-th molecule; d_{ij} is the dissimilarity of molecules *i* and *j*; *q*[0, ∞] is the order of diversity.

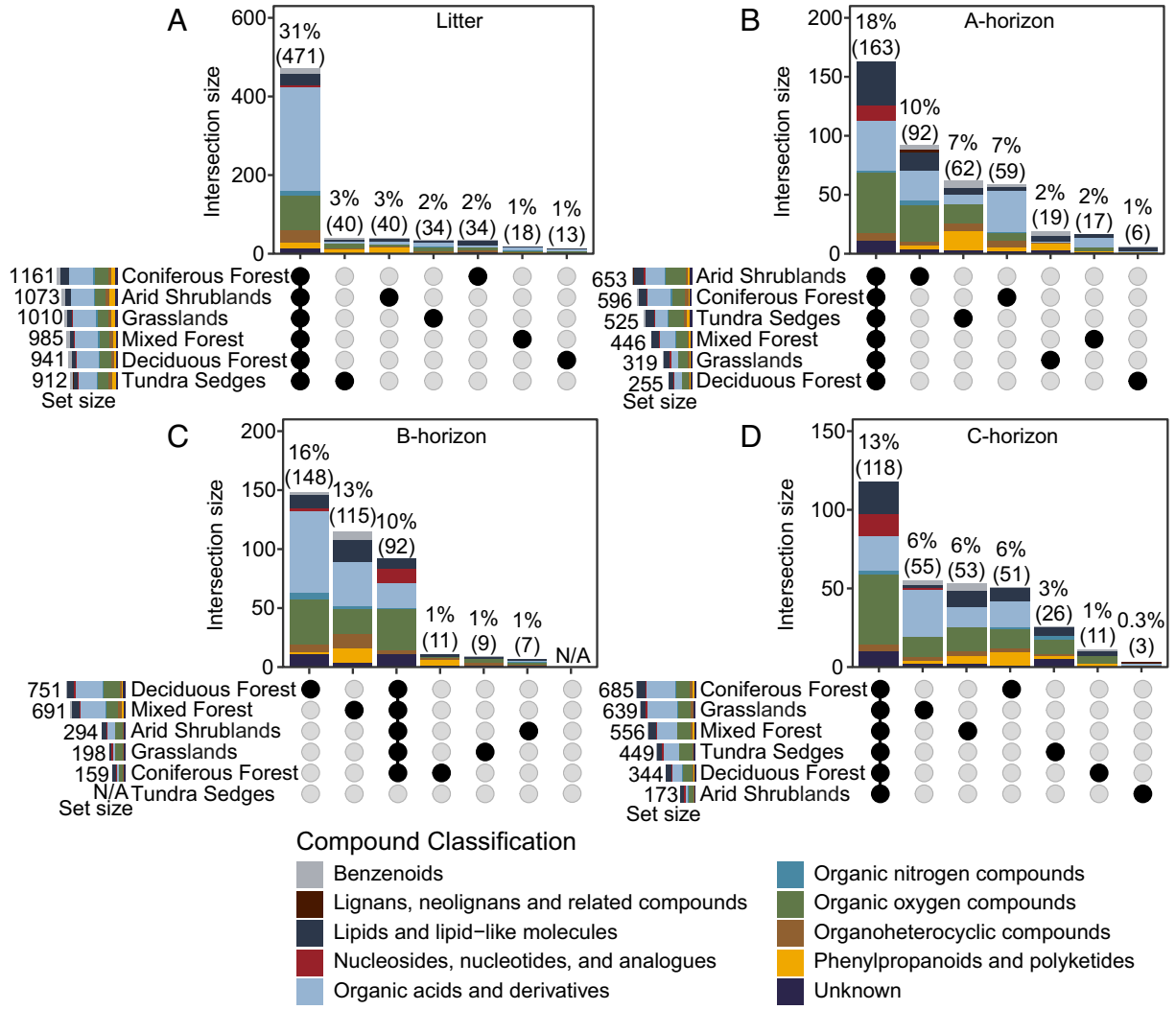

Fig. 1. Shared and unique hydrophilic compounds (HILIC LC–MS/MS) identified in each horizon. Both shared and unique compounds are displayed for the (*A*) litter layer, (*B*) A-horizon, (*C*) B-horizon, (*D*) C-horizon. Black dots under vertical bars indicate sets of ecosystems considered; either as individual ecosystems (single black dot) or all ecosystems (six black dots in the litter, A-, and C-horizons or five black dots in the B-horizon). The proportion of unique compounds, that occur only in a single ecosystem (single black dot), and shared compounds, those that are common across all ecosystems (all black dots shaded), is shown. The proportion of compounds that are either shared or unique is displayed above the vertical bars, with the number of compounds making up that proportion shown in parentheses below. The set sizes, or the total number of compounds identified for each ecosystem, are shown as horizontal bars. The identified compound classifications are reported by color within both the vertical and horizontal bars. Proportions and number of features common or unique of the sum of features in each horizon are displayed (proportions missing to 100% are features that are neither common nor unique; NA not available). Shared and unique features for hydrophobic compounds showed similar trends and are displayed in *SI Appendix*, Fig. S7.

layer, organic acids were the most represented class (56%) (Fig. 3) and included compounds such as gluconic acid, betaine, and di-peptide metabolites (e.g., Phe-Ala and Ile-Ala) (Dataset S2). The A-horizon had relatively fewer shared hydrophilic compounds than the litter horizon (–13%; Fig. 1). Hydrophilic compounds common to all A-, B-, and C-horizons were primarily oxygen-containing compounds, such as arabitol, glyceric acid, and N-acetyl-D-glucosamine (Fig. 3 and Dataset S1). Hydrophilic "unique" compounds (i.e., annotated molecular features that occurred in a single ecosystem or horizon) exhibited the greatest decline in the arid shrublands from the litter to the C-horizon (–45%, Fig. 2); the C-horizon contained only two compounds (D-gluconic acid and chelidonic acid) that were not present in the other horizons (Dataset S1). Interestingly, across all the six ecosystems, the litter layers were more compositionally similar to each other than to the C-horizons (Fig. 5), suggesting hydrophilic compounds become increasingly dissimilar across different ecosystems.

Hydrophobic molecular features displayed similar trends. Litter layers shared over 12 times the number of compounds across ecosystems as C-horizons (*SI Appendix*, Fig. S7) and were primarily organic acids and derivatives (41%) and benzenoids (12%) in A-horizons, and organic acids and derivatives and nucleosides/nucleotides/analogs in B- and C-horizons (Fig. 3 and Dataset S1). Relative to the hydrophilic compounds, hydrophobic compounds in the litter layer were more dissimilar than the C-horizon (*SI Appendix*, Fig. S9*B*), suggesting that this molecular suite of compounds becomes increasingly similar with depth across ecosystems.

Overall, SOM molecular richness (D_R) and abundance-based diversity ($D_{H,\ q=2}$) decreased with depth (Fig. 4 and *SI Appendix*, Figs. S5 and S6 and Tables S3 and S4), but trends differed among the six ecosystems. Depth was a significant predictor for abundance-based molecular diversity (hydrophilic compounds: $F_{1,5} = 9.17$, $P < 0.01$; hydrophobic compounds: $F_{1,5} = 13.11$, $P < 0.001$) and molecular richness (hydrophilic compounds: $F_{1,5} = 11.79$, $P < 0.0001$; hydrophobic compounds: $F_{1,5} = 18.99$, $P < 0.001$), but not for trait-based functional molecular diversity ($FD_{Rao(NOSC)}$) (*SI Appendix*, Figs. S5 and S6 and Tables S3 and S4). Across all

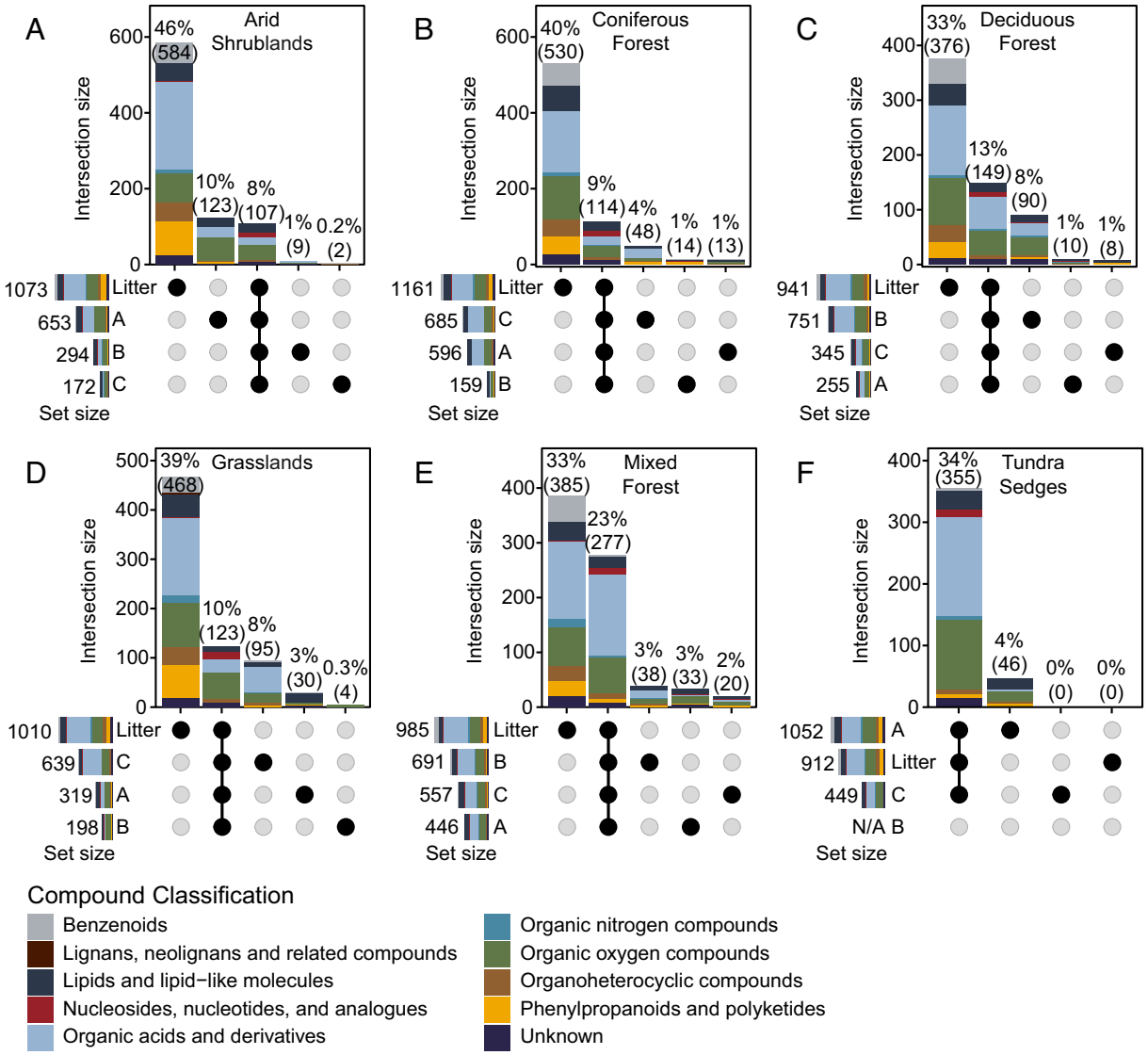

Fig. 2. Shared and unique hydrophobic compounds identified in each ecosystem type. Both shared and unique compounds are displayed for the litter, A-horizon, B-horizon, and C-horizon for the six ecosystem types: (*A*) arid shrubland, (*B*) coniferous forest, (*C*) deciduous forest, (*D*) grassland, (*E*) mixed forest, and (*F*) tundra sedges. Black dots under vertical bars indicate sets of horizons considered; either as individual horizons (single black dot) or all horizons (four black dots). The proportion of unique compounds, that occur only in a single horizon (single black dot), and shared compounds, those that are common across all horizons (all black dots shaded), is shown. The proportion of compounds that are either shared or unique is displayed above the vertical bars, with the number of compounds making up that proportion shown in parentheses below. The set sizes, or the total number of compounds identified for each horizon, are shown as horizontal bars. The identified compound classifications are reported by color within both the vertical and horizontal bars. The proportion of features common or unique of the sum of features in each ecosystem is displayed (proportions missing to 100% are features that are neither common nor unique; NA not available). Shared and unique features for hydrophobic compounds showed similar trends and are displayed in *SI Appendix*, Fig. S8.

ecosystems, molecular richness and abundance-based molecular diversity of hydrophilic compounds declined from the litter layer to the topsoil A-horizon (Fig. 4 and *SI Appendix*, Fig. S4 and Tables S3 and S4), with the greatest decline in deciduous forests (−79%, *P*-value < 0.05) and the smallest decline in arid shrublands (−48%, *P*-value < 0.05) (Fig. 4*A* and *SI Appendix*, Table S3). Although most of the ecosystems showed a decline in hydrophilic molecular richness throughout the soil profile, there were large increases in molecular richness from the B- to the C-horizon in the coniferous forests (+432%, *P*-value < 0.05) and the grasslands (364%, *P*-value <0.05) (Fig. 4*A*). The molecular richness of hydrophobic compounds declined from the litter to the A-horizon, with the greatest decline in arid shrublands (−84%, *P*-value < 0.05) and the smallest declines in mixed forest (−60%, *P*-value < 0.05) and tundra sedge ecosystems (−61%, *P*-value < 0.05) (*SI Appendix*, Fig. S5 and Table S4). Trends in abundance-based molecular diversity mimicked molecular richness (*SI Appendix*, Fig. S5 and Tables S3 and S4), but the changes across depth and ecosystems were less pronounced.

The decline in molecular diversity with depth coincided with an increase in absorbance of the aliphatic C–H bonding energy detected in ATR-FTIR spectra (*SI Appendix*, Fig. S2), indicating a greater abundance of microbial-derived compounds with increasingly negative NOSC values (41). Similarly, the amide N–H and aromatic C=C bonds (−1,720 to −1,700 cm^{-1}) shifted toward the amide C=O bonding region (−1,660 to −1,630 cm^{-1}) with depth. ATR-FTIR data showed a greater shoulder at 1,710 cm^{-1} indicative of carboxyl C–O which is consistent with NEXAFS spectra of bulk soils from selected locations exhibiting an increase in the ratio of carboxylic to aromatic C functional groups with depth (*SI Appendix*, Fig. S3). These complementary analytical approaches suggest that declines in molecular richness with depth are driven by increasing contributions of microbial-derived products.

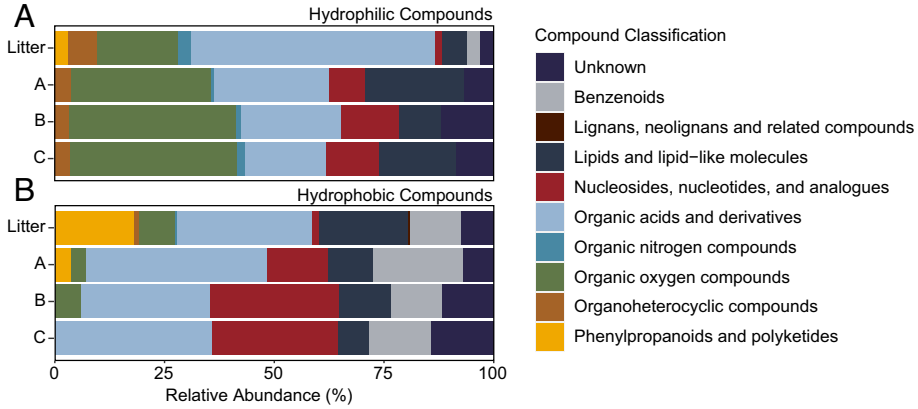

Fig. 3. Compound classifications of identified molecules that occurred in individual horizons (litter, A, B, C) from each are identified for hydrophilic compounds (HILIC LC–MS/MS) (*A*), and the hydrophobic compounds (C18 LC–MS/MS) (*B*). Compounds are classified by "superclass" using the GNPS data analysis tool for LCMS/MS data. Compounds that were positively identified but not annotated are designated "Unknown." Note, the number of annotated compounds used to calculate relative abundances generally decreased from the litter to the C-horizons (e.g., we identified 471 hydrophilic compounds in the litter horizon and 118 compounds in the C-horizon).

Molecular Diversity in Relation to Ecosystem Type. The molecular composition of plant litter inputs varied by ecosystem type. Plant litter C:N ratios were highest in coniferous forests (71.9 ± 9.3, *P*-value < 0.05) and lowest in grasslands (26.4 ± 3.0, *P*-value < 0.05) (*SI Appendix*, Table S2). Aromatic C functional groups were 6% lower in tundra sedges than that in coniferous forests or arid shrublands (NEXAFS spectra; *SI Appendix*, Fig. S3). Tundra sedges also had more pronounced peaks in the aliphatic

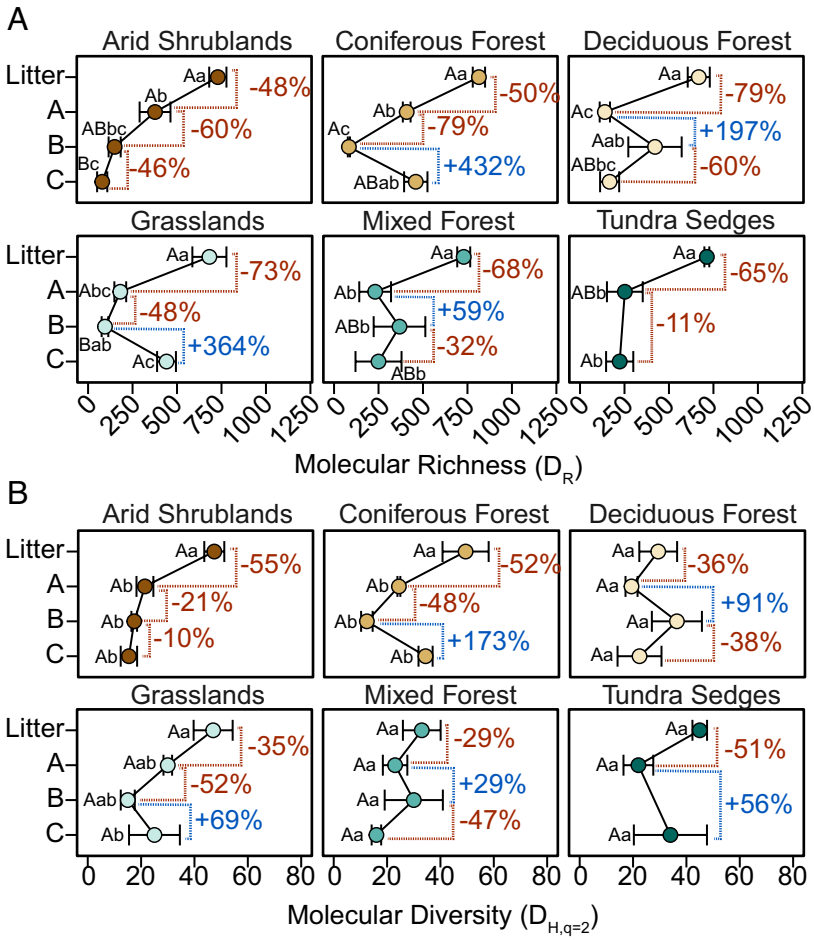

Fig. 4. Molecular richness (*A*) and molecular diversity (*B*) of the litter, A-, B-, and C- horizons of hydrophilic compounds (HILIC LC–MS/MS) within six ecosystems. Percent differences between the litter and A-horizons, A- and B-horizons, and B- and C-horizons show the percent increase (blue) or decrease (red) in molecular richness and diversity. Significance differences are displayed with capital and lowercase letters. Capital letters indicate differences within a given horizon (A, B, or C or litter) across ecosystems. Ecosystems that do not share the same capital letter are significantly different. Lowercase letters indicate differences between horizons within a single vegetation class. Horizons that do not share the same lowercase letters are significantly different. Differences of means were determined with mixed-effects models using a Bonferroni correction for 3 or 15 tests. Trends for the hydrophobic compounds were similar, but weaker, than those of the hydrophilic compounds (*SI Appendix*, Fig. S6).

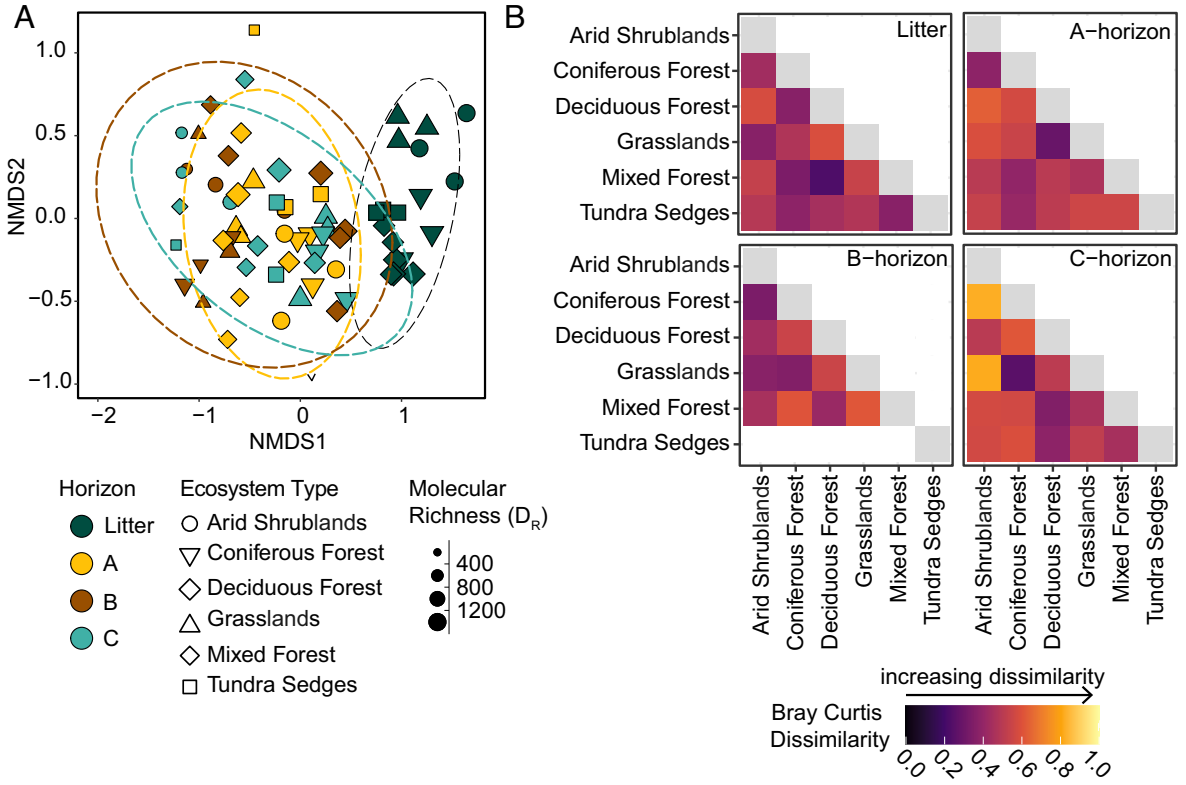

Fig. 5. Distribution of hydrophilic compounds (HILIC LC–MS/MS) using NMDS ordination (stress = 0.14) (*A*) and Bray–Curtis Dissimilarity matrices (*B*). Potential predictor variables included as vectors in the NMDS ordination were mean annual temperature (°C); mean annual precipitation (mm); latitude; longitude; elevation (m); depth (m); concentrations of SOC, total nitrogen, and hydroxylamine extractable iron (mg g^{-1} soil); clay content; pH; and gravimetric moisture content. No vectors were found to significantly explain ordination variance (*P*-value < 0.05). PERMANOVA analysis show that ecosystem type explained 10% ($P < 0.001$) and horizon explained 21% ($P < 0.001$) of the variance in dissimilarity of LC-MS/MS SOM samples. The interaction of ecosystem type and horizon explained 24% of the variance in dissimilarity. Average Bray–Curtis distance matrices (used in the NMDS ordination) show the dissimilarity of ecosystems grouped by horizon. Trends were generally similar of hydrophobic compounds; however, the litter showed greater dissimilarity compared to the hydrophilic compounds (*SI Appendix*, Fig. S9).

C–H bonding region than that of the other ecosystems (FTIR spectra; *SI Appendix*, Fig. S2). Grasslands had more pronounced amide N–H and aromatic C=C features than those of the other ecosystems (*SI Appendix*, Fig. S2). Despite initial differences in litter composition, models (including both ecosystem and horizon as fixed effects) did not significantly improve the null models (including only horizon as a fixed effect) predicting molecular richness, abundance-based molecular diversity, or trait-based functional diversity metrics (*SI Appendix*, Tables S7 and S8).

The largest differences in molecular richness across ecosystems occurred in the litter layer (Fig. 4 and *SI Appendix*, Fig. S6 and Tables S3 and S4). For example, within the litter layer, the molecular richness of hydrophilic compounds was lowest in deciduous forests (671 compounds ± 62) and highest in coniferous forests (818 ± 35) (Fig. 4 and *SI Appendix*, Table S3). The molecular richness of hydrophobic compounds was lowest in tundra sedges (263 ± 53, *P*-value < 0.05) and grasslands (277 ± 24, *P*-value < 0.05) and highest in coniferous forests (485 ± 13) (*SI Appendix*, Fig. S5 and Table S4). Abundance-based and trait-based functional molecular diversity indices of hydrophilic and hydrophobic compounds did not vary significantly across soil horizons in any ecosystem (Fig. 4 and *SI Appendix*, Fig. S6 and Tables S3 and S4).

The relative abundance of hydrophilic compounds present in all the four horizons of the soil profile ranged from 8 to 34% in the six ecosystems (Fig. 2). Tundra sedges shared the greatest number of hydrophilic compounds across the soil profile (34% shared across three soil horizons, Fig. 2), whereas hydrophilic SOM in arid shrublands, coniferous forests, and grasses shared the fewest number of molecular features (8, 9, and 10%, respectively, averaged across four soil horizons) (Fig. 2). Interestingly, SOM in the tundra sedges litter layer also contained the greatest number of unique hydrophilic compounds (40 compounds, or 3% of the molecular features), which were not present in the corresponding horizons of the other five ecosystems (Fig. 1*A*). Grasslands contained more unique hydrophilic compounds in the C-horizon than those of the other ecosystems (55 compounds, 6%), while arid shrublands contained the fewest (3 compounds, 0.3%) (Fig. 1*D*).

Arid shrublands had the fewest number of shared hydrophobic compounds in all horizons (13 compounds, 2%), which were primarily di-peptide organic acids and nucleosides such as 2′-Deoxyadenosine. The tundra sedges shared 21% of hydrophobic compounds (110 compounds) across the three soil horizons (*SI Appendix*, Fig. S8), which were also primarily di-peptide organic acids. In contrast, the greatest number of unique hydrophobic compounds were found in the litter layer of coniferous forests and arid shrublands (40 compounds, 5% each) and the C-horizon of mixed forests (49 compounds, 26%) (*SI Appendix*, Fig. S7).

Molecular Diversity in Relation to Environmental and Soil Properties. Molecular dissimilarity of SOM was largely driven by depth (Fig. 5 and *SI Appendix*, Fig. S9), and specifically by differences in litter-derived SOM, which clearly separated in ordination space from the other soil horizons [nonmetricdimensional scaling (NMDS) ordination of hydrophilic compounds based on Bray–Curtis dissimilarity; Fig. 5*A*]. The C:N ratio of the litter layer was the strongest predictor of molecular richness for hydrophilic ($R^2 = 0.14$, *P*-value = 0.09) and

hydrophobic compounds (R^2 = 0.26, *P*-value = 0.02) (*SI Appendix*, Tables S5 and S6); however, this relationship did not continue in the mineral soil horizons. The interaction of ecosystem and horizon explained 24% ($P < 0.001$; PERMANOVA) of the variance of hydrophilic compounds, with horizon and ecosystem accounting for 21% ($P < 0.001$; PERMANOVA) and 10% ($P < 0.001$; PERMANOVA) of the variance, respectively (Fig. 5*A*). NMDS ordination of hydrophobic compounds showed a similar clustering pattern to hydrophilic compounds, with ecosystem explaining 17% of the variance (*P*-value < 0.001; PERMANOVA) and horizon explaining an additional 17% of the variance (*P*-value < 0.001; PERMANOVA), and their interaction accounting for 20% of the variance ($P < 0.001$; PERMANOVA) (*SI Appendix*, Fig. S9). None of the potential explanatory variables included in the NMDS ordination were significant (*P*-values > 0.05). These included climatic variables (mean annual temperature and precipitation), environmental location characteristics (latitude, longitude, elevation), soil depth, concentrations of SOC, total nitrogen, hydroxylamine-extractable iron, clay content, soil pH, and gravimetric water content. These results are supported by linear regressions, which showed that no environmental or soil variables were correlated with molecular diversity indices of hydrophilic or hydrophobic compounds in the mineral soil horizons (*SI Appendix*, Tables S5 and S6).

Discussion

Molecular Richness and Diversity Are Driven by Microbial Transformation Rather Than Initial Plant Litter Chemistry. Despite large differences in the molecular composition and C:N ratios of litter inputs, the absolute number of molecules and both abundance and trait-based diversity indices did not differ among ecosystems. Subsoil horizons had a stronger molecular signature of microbial-derived OM than that of topsoil horizons (based on FTIR and LC–MS/MS spectra), which coincided with lower molecular α-diversity in the subsoil (Fig. 4 and *SI Appendix*, Figs. S4–S6). The decrease in molecular α-diversity with depth coincides with a substantial decrease in molecularly unique compounds as litter-derived OM is incorporated into mineral horizons (Fig. 2 and *SI Appendix*, Fig. S8). The few shared compounds in the subsoils across ecosystems were predominantly organic acids and derivatives, oxygen-containing compounds, and nucleosides/nucleotides/analogs, most of which were indicative of microbial metabolism such as 2′-Deoxyadenosine (purine nucleoside component of DNA), N-acetyl-D-glucosamine (component of bacterial cell walls), and arabitol (a microbial metabolite) (42). These compounds indicate that increased microbial transformation of plant litter imparts a microbial signature in SOM profiles (43–47). Our findings are consistent with previous results, where microbial decomposition transforms molecularly diverse plant litter inputs to a pool of molecularly similar, microbially derived compounds (9, 48).

Molecular Diversity of SOM Changed by Microbial Transformation. Microbial decomposition reduced the number and diversity of molecules cycling downward through the soil profile (Fig. 4 and *SI Appendix*, Figs. S4–S6), although site-specific soil properties moderated these patterns. For example, the molecular richness and diversity of grassland and coniferous forest C-horizons was lower than that of the litter layer, but higher than that of the A- or B-horizons. This pattern could be due to higher root C inputs at depth, downward transport of SOM, or physical translocation due to burrowing animals such as insects or earthworms (49).

After accounting for the oxidation state and molecular weight of each compound, we found that functional molecular diversity was not influenced by microbial decomposition (*SI Appendix*, Figs. S4 and S5 and Tables S3 and S4). As microorganisms convert plant-derived molecules (with varying NOSC values) into microbial-derived products, they generate products that are similarly oxidized throughout the soil profile. In contrast, marine SOM was composed of molecularly dissimilar compounds across an aquatic degradation gradient, despite the fragmented molecules displaying high similarity across sites (12). Using a laboratory incubation experiment, others showed that the trait-based functional diversity, based on NOSC values, of marine SOM declined with increasing degradation time (27); both the abundance and molecular richness of marine SOM increased, but began to decline after 100 d of incubation, ultimately returning to starting values after 1,000 d (27). These patterns could result from a lack of new C inputs to the experiment. Our observational findings, based on natural soil and litter samples, demonstrate that continuous organic C inputs (i.e., from roots or downward translocation) led to functionally similar molecules with increased decomposition (*SI Appendix*, Figs. S4 and S5).

Microbial Transformation Results in Convergent α-Diversity and Divergent β-Diversity. The process of hydrolytic and oxidative enzymatic decomposition by microorganisms generates SOM that becomes increasingly similar relative to the molecular diversity of plant litter inputs (Figs. 1, 2, and 4 and *SI Appendix*, Figs. S4–S9). Through microbial consumption of plant-derived organic C, microbes sustain catabolic and anabolic metabolism, converting SOC to respired CO_2 (or CH_4) or microbial biomass, which may contribute to the SOM pool (7). This decomposition process, referred to as the microbial funnel, suggests that the bulk of the SOM is composed of microbial metabolic by-products (1, 7, 50–52). These by-products include cellular waste, exudates, primary and secondary metabolites, and cell wall fragments (44, 47, 52, 53). Despite differing ecosystem conditions and plant litter inputs, the continued turnover and accrual of microbial by-products leads to a lower number of molecules that are composed of a similarly diverse array of molecules over time (*SI Appendix*, Figs. S4 and S5).

Even though shared compounds across sites were indicative of microbial decomposition, the total number of shared hydrophilic and hydrophilic compounds across ecosystems was much lower in the subsoil compared to the initial litter samples. The low number of shared hydrophilic compounds in the subsoils was associated with a higher degree of dissimilarity across sites (Fig. 5*B*), suggesting an overall increase in β-diversity and decrease in α-diversity with depth (Fig. 6). In comparison, hydrophobic compounds had a higher β-diversity value (based on the Bray–Curtis Dissimilarity Index) with depth (*SI Appendix*, Fig. S9) despite sharing 10% fewer compounds across ecosystems than that of the hydrophilic compounds (Fig. 1 and *SI Appendix*, Fig. S7). The difference in Bray–Curtis β-diversity indices between the hydrophilic and hydrophobic compounds could be due to their polar or nonpolar nature—which dictates how mobile the compounds are within the soil column and thus how similar the compounds are to one another within the differing soil horizons (8).

Together, these results suggest that microbial turnover of SOM reduces overall molecular diversity within each ecosystem, although the individual ecosystems yield site-specific compounds that often exhibit divergent chemistries with depth. Site-specific molecular dissimilarity could stem from differences in microbial community structure and function (54, 55) or mineralogy (56, 57). For example, microbial communities that assemble in different ecosystems may produce unique metabolic profiles based on soil conditions,

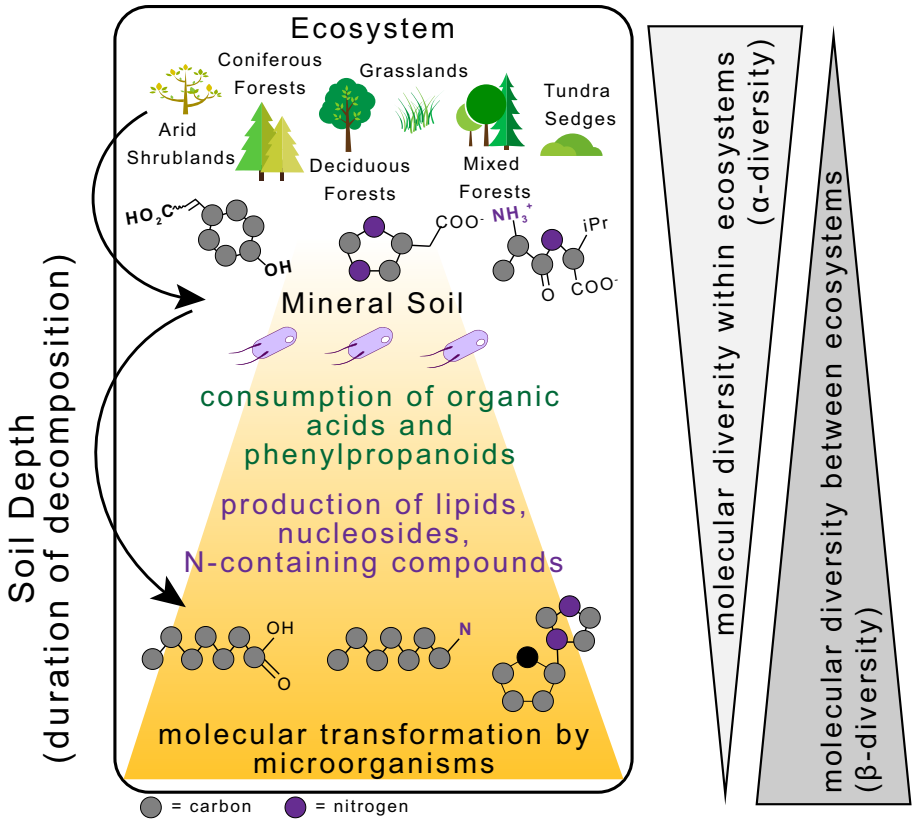

Fig. 6. Conceptual diagram of changes in α- and β-diversity with depth and changes in shared SOM composition across ecosystems. α-diversity decreases as plant litter inputs are microbially transformed and incorporated into mineral soil horizons (Fig. 1 and *SI Appendix*, Figs. S4–S6). Decreases in organic acids and phenylpropanoids occur in each ecosystem while nucleosides and nitrogen-bearing compounds increase. Microbial transformation of plant litter yields high β-diversity in the subsoils, turning SOM from plant litter with low site specificity into SOM that has high site specificity.

such as pH and oxygen availability, while site-specific minerals may differentially adsorb molecules based on molecular weight, heteroatom compatibility, or degree of carbon saturation (58). We further speculate that resulting SOM pools with low α-diversity found in deeper soil horizons, regardless of ecosystem type, are less energetically favorable for continued microbial catabolism and thus would have a greater persistence in soils (1).

While plant diversity is commonly linked with SOC sequestration potential and microbial diversity (32, 59, 60), our data show that ecosystem type did not improve model predictions of molecular α-diversity but did explain greater β-diversity with depth. In contrast, soil depth had a significantly stronger correlation with molecular α-diversity, than with molecular β-diversity. Thus, molecular α-diversity is likely governed by microbial degradation, and molecular β-diversity may be more influenced by site-specific characteristics, such as soil pH, microbial community, soil texture, and mineralogy. These site-specific soil properties could potentially increase β-diversity across ecosystems by dictating which compounds are utilized by microorganisms, occluded within aggregates, or adsorbed to mineral surfaces. Therefore, we conclude that microbial degradation is the single greatest factor controlling molecular α-diversity in a given soil, while molecular β-diversity is likely governed by numerous soil properties across ecosystems.

The observed convergence of molecular α-diversity in soil aligns with reductions in microbial community diversity generally reported with depth (61). Microbial and molecular α-diversity are strongly linked, yet molecular diversity has been shown to have a stronger effect on shaping the microbial community structure than vice versa (13). Thus, the convergence of molecular α-diversity along microbial degradation gradients may contribute to the general decline in microbial diversity. The observed divergence of molecular β-diversity of hydrophilic compounds suggests that site-specific compounds predominantly occur in the subsoils where organic matter has undergone more extensive decomposition (Fig. 6). In contrast, the apparent increase in molecular similarity with depth of hydrophobic compounds suggests that more unique nonpolar compounds are more resistant to transport down the soil profile. Dissolved organic matter analysis from across the North Atlantic and Southern Ocean showed that molecular α-diversity increased while β-diversity decreased across a degradation gradient (12). This study defined their degradation gradient as a longitudinal gradient from a lake to the deep ocean, whereas our study used soil depth as degradation gradient for each individual ecosystem. Our study therefore permitted the assessment of molecular diversity along degradation gradients both within individual ecosystems (α-diversity) and between contrasting ecosystems (β-diversity).

Advancements in technology continue to improve our ability to explore the molecular complexity of SOM. Here, we show that the molecular identity of SOM is critical to predicting how molecular diversity drives soil carbon cycling. Overall, we suggest that molecular α-diversity converges with depth during microbial decomposition of plant litter, but molecular β-diversity diverges with microbial decomposition across ecosystems. In contrast to our expectations, molecular properties based on NOSC and molecular weight were not suitable metrics for characterizing functional molecular diversity as a result of microbial decomposition. While trait-based functional diversity metrics are commonly calculated using alternative analytical methods, such as Fourier-transform-ion cyclotron resonance-mass spectrometry (FT-ICR-MS) (27, 28),

these metrics may not be suitable using LC–MS/MS. Assessing molecular diversity using the number of molecules and identifying shared molecules across groups, while also incorporating molecular traits, may further improve our understanding of how SOM molecular diversity shapes SOC mineralization rates and/or adsorption to mineral surfaces. Our work provides a framework for future insights into SOM molecular diversity and ecosystem function.

Materials and Methods

Soil Sampling. Soil and litter samples were collected in July of 2019 from locations across the United States from six distinctly different ecosystems (*SI Appendix*, Fig. S1). Soils sampled by horizon to the average depth of 1 m from each location were transported on ice to Cornell University where they were sieved at <2 mm and air-dried prior to analysis. Locations were grouped by region, ecosystem type, based on the predominant vegetation class, and soil order (*SI Appendix*, Table S1).

Soil Characterization. Particle size distribution was assessed by mechanical separation using the hydrometer method. Soil pH was measured in a 10-mM calcium chloride solution in a 1:2.5 (soil:solution, w/v) ratio. Poorly crystalline iron, aluminum, and manganese were determined using a hydroxylamine hydrochloric acid extraction (62); 0.33 g of soil and 10 mL of 0.25 M hydroxylamine HCl solution were shaken for 16 h and filtered at 0.20 μm. Elemental contents were quantified in the extracts using inductively coupled plasma mass spectrometry (Spectro Ametek). Dissolved organic carbon was analyzed using a Shimadzu TOC-L analyzer (Shimadzu Scientific Instruments, Inc.) after shaking 1 g of air-dried soil with 10 mL deionized water and then filtering at 0.20 μm. Total SOC and total nitrogen were measured with a Thermo Delta V isotope ratio mass spectrometer interfaced to an NC2500 elemental analyzer (Thermo Fisher Scientific Corp.). Mean annual temperature (°C), mean annual precipitation, and aridity were calculated for the latitude/longitude coordinates for the sampled locations using values from the WorldClim 2.0 datatset (63).

NEXAFS. Carbon functional groups of bulk organic carbon in soil and litter samples were determined using C 1(s) near-edge X-ray absorption fine structure (NEXAFS) spectroscopy at the Canadian Light Source in Saskatoon, Canada. Soil and plant litter samples from full profiles from each grouping (e.g., grassland, coniferous forests, arid shrubland, deciduous forests, mixed forests, and tundra tussock sedges and forbes) were ball-milled, slurried in deionized water, and pipetted onto clean silicon foils. Upon drying, C NEXAFS spectra were obtained with the spherical grating monochromator beamline 11ID-1 (64). Step scan mode (0.25 eV steps from 270 to 320 eV) was used to minimize X-ray damage. A dwell time of 20 ms was used between scans. Individual spectra were collected at new locations on each sample for a total of 20 to 30 scans. The beamline exit slit was set at 25 mm, and the fluorescence yield data were collected using a two-stage microchannel plate detector. The resulting spectra were averaged for each sample and the averaged spectrum was then baseline normalized to zero and then normalized using the beamline photon flux (I_0) from a separate Au reference foil. Each spectrum was calibrated to the carboxylic acid peak (288.5 eV) of a citric acid standard. Preedge (270 to 278 eV) and postedge (310 to 320 eV) and E_0 (290 eV) values were used to perform an edge-step normalization. Peak deconvolution was conducted in Athena [Demeter version 0.9.25, 2006 to 2016; (65)] to determine the relative abundances of functional groups, namely carboxylic and amine C, aromatic C, and aliphatic C (66, 67). Gaussian peak positions, their full-width at half-maximum, and the arc tangent function were fixed. Peak magnitude was set to vary freely during the fitting process. Parameters were adjusted until optimal fits for each spectrum were achieved, and all spectra were fitted with these final parameters.

FTIR Spectroscopy. Fourier transform infrared (FTIR) spectra of water-extractable organic matter (<0.2 μm) from soil and plant litter samples were obtained using a Bruker Vertex 70 spectrometer (Bruker Optics Inc.) equipped with a Pike GladiATR accessory (Pike Technologies) and using a single-reflection diamond internal reflection element. The spectra were collected from 4,500 to 150 cm^{-1} with a resolution of 4 cm^{-1} and averaged from 60 scans per sample. OPUS 7.2 software (Bruker Optics Inc.) was used to conduct baseline correction and normalization before exporting spectra for interpretation.

Metabolite Extraction. Soil or litter material was weighed out at 2.6 to 5.8 mg of water-extractable C equivalent and then resuspended in 4 mL LCMS-grade water per gram of soil or 5 mL LCMS-grade water per gram of leaf litter. Samples were vortexed for 10 s, then bath sonicated in an ice water bath for 1 h. The samples were incubated for an additional hour in an ice water bath and then shaken overnight at 4 °C at 200 rpm on an orbital shaker. Slurries were centrifuged at 3,000 × g for 10 min at 10 °C and the supernatants were dried in a new tube by lyophilization (Labconco FreeZone 12L). The soil pellets were resuspended in 5 mL 2:1 v:v chloroform:methanol. The samples were vortexed for 10 s, sonicated for 30 min in an ice water bath, and then shaken at 4 °C for 1 h at 200 rpm on an orbital shaker. The samples were centrifuged (10,000 rcf, 10 min, 10 °C) and the supernatants were dried under vacuum centrifugation (Thermo Savant Speed Vac). Dried water and organic solvent extracts were both resuspended in 1 mL methanol, vortexed 10 s, bath sonicated 10 min in ice water, and then combined into one tube; vortexing and sonication was repeated, and then the samples were centrifuged at 10,000 rcf for 5 min at 10 °C. The supernatants of the combined extracts were dried by vacuum concentration (Thermo Speed-Vac) and then resuspended in 150 μL methanol containing internal standards (Dataset S2); the final extracts were vortexed 2 × 10 s, bath sonicated in ice water for 15 min, and then centrifuged (10,000 rcf, 5 min, 10C) and filtered (0.2 μM) to remove insoluble materials; the filtrates were collected for LC–MS/MS.

LC–MS/MS. Metabolites from extracts were separated using reverse phase on an Agilent 1290 LC stack connected to a Q Exactiv Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Samples were injected twice, once on reverse-phase LC–MS/MS (Agilent ZORBAX RRHD Eclipse Plus C18, 95Å, 2.1 × 50 mm, 1.8 μm column) and again on hydrophilic interaction LC–MS/MS (Agilent InfinityLab Poroshell 120 HILIC-Z, 2.1 × 150 mm, 2.7 μm column). LC–MS/MS parameters are described in Dataset S2. Briefly, after separation, the top two most intense precursor ions (not previously selected within 7 s) were selected for fragmentation at stepped collision energies of 10, 20, 40 eV. Sample injections were randomized and both internal and external standards were used for quality control purposes. Chromatographic features and MS/MS spectra were extracted using MZmine 2 prior to annotation via spectral matching using GNPS libraries (68, 69) and the Feature-Based Molecular Networking workflow (70, 71). An MZmine workflow was used to generate a list of features (mzrt values obtained from extracted ion chromatograms containing chromatographic peaks within a narrow m/z range) and filtered to remove isotopes. For each feature, the most intense fragmentation spectrum was uploaded to GNPS. When a sample mass spectrum matches the one deposited within the GNPS database, a putative identification is made. Library hits were filtered based on accurate mass <15 ppm for negative mode and <5 ppm for positive mode, cosine score >0.70, and number of matching ions = 3. In addition, features were filtered to meet the following requirements: retention time greater than 1 min and intensities that were 50 times greater than the extraction controls. Only the top one hit from the GNPS networking library was selected to remain in the dataset. Features were also dropped when all peak heights across samples were less than 1,000,000 intensity. For further statistical processing, feature table intensities were normalized by NPOC (total mg of C) in the soil/litter extracted.

Molecular Diversity Calculations. Molecular diversity of SOM extracted from soil and litter samples was calculated using a variety of classic and functional diversity indices. We used classic functional diversity indices as measures of molecular α-diversity, meaning the diversity of compounds within one sample or ecosystem. Functional diversity calculations and distant matrixes were used as measures of molecular β-diversity, or the similarity of compounds across sites sites/ecosystems. Selection of diversity indices was carefully considered, especially since mass spectrometry data are uniquely different from classic ecology datasets often used to calculate diversity (e.g., species count or operational taxonomic unit data); hence, a full discussion of diversity indices for mass spectrometry data can be found in supplemental information and brief descriptions are found in Table 1.

We calculated i) molecular richness (D_R) using the sum of identified molecular features in each sample (*SI Appendix*, Tables S3 and S4), ii) abundance-based molecular diversity using Hill Numbers (72) for an evenly weighted pool of molecules ($D_{H,\,q=1}$) and for a pool of molecules that favors rare individuals ($D_{H,\,q=2}$), and iii) trait-based functional molecular diversity using Rao's quadratic entropy (27, 34) with either NOSC ($D_{Rao(NOSC)}$) or molecular weight ($D_{Rao(MW)}$) as the dissimilatory

molecular property (Table 1). All indices are reported in *SI Appendix*, but in the main text, we limit our discussion to molecular richness (D_R), molecular diversity when $q = 2$ ($D_{H, q=2}$), and functional molecular diversity using NOSC ($D_{Rao(NOSC)}$).

Molecular richness and abundance-based molecular diversity were calculated using peak intensities as presence/absence and relative abundance, respectively, of unique annotated metabolite features. The subset of extracted LC–MS/MS features that were annotated to a chemical identifier and associated molecular formula based on GNPS spectral matching were used for richness or diversity measures. The molecular formula for the top GNPS match was used to calculate molecular properties based for functional molecular diversity including molecular weight (amu) and the NOSC (41).

Statistical Analysis. Soil biogeochemical variables were determined to be significantly different by testing the means using a mixed-effects model, where horizon (litter, A, B, C) and ecosystem type were calculated as fixed effects and the location as a random effect. Variables were log transformed when the model residuals were not normally distributed. Variable means were contrasted using estimated marginal means with a Bonferroni correction factor using the emmeans R package (73) and compact letter displays. Variables were contrasted to first test whether individual horizons differed across ecosystems (capital letters) or within individual ecosystems to test differences in horizons (lowercase letters). Groups that do not share the same letters are considered significantly different (*SI Appendix*, Tables S2–S4). To identify predictors of molecular diversity, linear regressions were performed on horizon subsets of the data (*SI Appendix*, Tables S5 and S6). To determine if depth predicts molecular diversity, we used mixed-effects models for each diversity index which included depth (m) and ecosystem type as fixed effects and sample location as a random effect *SI Appendix*, Tables S7 and S8. Mixed-effects models were also conducted to determine the effect of ecosystem type on molecular diversity indices, where null models included depth (m) as a fixed effect and sample location as a random effect and were compared with ANOVA analysis to models that added ecosystem type as a fixed effect (*SI Appendix*, Tables S9 and S10). To identify dissimilarities of LC–MS/MS features across ecosystems, we performed a nonmetric dimensional scaling (NMDS) ordination using a Bray–Curtis distance matrix for relative abundances of LC–MS/MS features. Vectors of environmental variables were determined using the vegan package in Rstudio. Variables included as vectors were mean annual temperature (°C), mean annual precipitation (mm), latitude, longitude, elevation (m), depth (m), SOC, total nitrogen, clay, hydroxylamine extractable iron (mg g $soil^{-1}$), pH, gravimetric moisture content (GWC), NOSC, and molecular weight (MW). A PERMANOVA analysis was conducted to identify the extent to which ecosystem type explains the variance in the Bray–Curtis distance matrix with ecosystem type and horizon as factors and sample location as strata (74). Mean Bray–Curtis dissimilarity values for each ecosystem by horizon were also calculated using identified LC–MS/MS features for both the hydrophilic and hydrophobic compounds.

Data, Materials, and Software Availability. LC–MS/MS peak height tables and all associated metadata are available through the following GNPS job links: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=98e3c66affda45588ceb8d4b4534a4f9 (HILICZ positive mode), https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a360bd4d466948dd8929aa9bf803a18e (HILICZ negative mode), https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7988718f02d94fd7a10b3039de2b9e37 (C18 positive mode), https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=cb8d264c7a5d4610a93682365a9323ae (C18 negative mode). All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. Funding for this project was provided by the U.S. Department of Energy (DOE), Office of Biological & Environmental Research Genomic Science Program under award number DE-SC0016364, and a Schmittau-Novak Small Grant through Cornell University. T.R.N., B.P.B., and S.M.K. were supported by the ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Science Focus Area Program at Lawrence Berkeley National Laboratory supported by the U.S. DOE, Office of Science, Office of Biological & Environmental Research, under contract DE-AC02-05CH1123, and used resources of the National Energy Research Scientific Computing Center, a DOE of Science User Facility operated under contract number DE-AC02-05CH11231. NEXAFS analyses were done at the Canadian Light Source (user proposal 012307), and the help by T. Regier and J. Dynes is much appreciated. Attenuated total reflectance-Fourier transform infrared analyses were done with the support of C.E. Martínez with help from B. Azimzadeh. We would like to thank C. Levitt for his work on this project, as well as E. Mudrak for her statistical guidance.

Author affiliations: [a]Soil and Crop Sciences, School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14850; [b]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Metabolomics Technology Group Joint Genome Institute, Department of Energy, Walnut Creek, CA 94598; [d]Department of Global Development, Cornell University, Ithaca, NY 14850; [e]Cornell Institute for Digital Agriculture, Cornell University, Ithaca, NY 14850; and [f]Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY 14850

1. J. Lehmann *et al.*, Persistence of soil organic carbon caused by functional complexity. *Nat. Geosci.* **13**, 529–534 (2020).
2. M. Allen *et al.*, Technical Summary: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Intergovernmental Panel on Climate Change (2019).
3. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
4. R. Baran *et al.*, Exometabolite niche partitioning among sympatric soil bacteria. *Nat. Commun.* **6**, 8289 (2015).
5. A. Canarini, C. Kaiser, A. Merchant, A. Richter, W. Wanek, Root exudation of primary metabolites: Mechanisms and their roles in plant responses to environmental stimuli. *Front. Plant Sci.* **10**, 1–19 (2019).
6. M. Chomel *et al.*, Plant secondary metabolites: A key driver of litter decomposition and soil nutrient cycling. *J. Ecol.* **104**, 1527–1541 (2016).
7. C. Liang, J. P. Schimel, J. D. Jastrow, The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2**, 1–6 (2017).
8. K. Kaiser, K. Kalbitz, Cycling downwards - dissolved organic matter in soils. *Soil Biol. Biochem.* **52**, 29–32 (2012).
9. V. N. Roth, T. Dittmar, R. Gaupp, G. Gleixner, The molecular composition of dissolved organic matter in forest soils as a function of pH and temperature. *PLoS One* **10**, 1–23 (2015).
10. M. W. I. Schmidt *et al.*, Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
11. T. Dittmar, *Reasons Behind the Long-Term Stability of Dissolved Organic Matter* (Elsevier Inc., ed. 2, 2014).
12. M. Zark, T. Dittmar, Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 1–8 (2018).
13. A. J. Tanentzap *et al.*, Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24689–24695 (2019).
14. V. N. Roth *et al.*, Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).
15. K. M. DeAngelis *et al.*, Long-term forest soil warming alters microbial communities in temperate forest soils. *Front. Microbiol.* **6**, 1–13 (2015).
16. S. C. Córdova *et al.*, Plant litter quality affects the accumulation rate, composition, and stability of mineral-associated soil organic matter. *Soil Biol. Biochem.* **125**, 115–124 (2018).
17. M. F. Cotrufo, M. D. Wallenstein, C. M. Boot, K. Denef, E. Paul, The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: Do labile plant inputs form stable soil organic matter? *Glob. Chang. Biol.* **19**, 988–995 (2013).
18. K. M. Geyer, P. Dijkstra, R. Sinsabaugh, S. D. Frey, Clarifying the interpretation of carbon use efficiency in soil through methods comparison. *Soil Biol. Biochem.* **128**, 79–88 (2019).
19. K. Geyer, J. Schnecker, A. S. Grandy, A. Richter, S. Frey, Assessing microbial residues in soil as a potential carbon sink and moderator of carbon use efficiency. *Biogeochemistry* **151**, 237–249 (2020).
20. E. G. Jobbágy, R. B. Jackson, The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
21. D. L. Jones, C. Nguyen, R. D. Finlay, Carbon flow in the rhizosphere: Carbon trading at the soil-root interface. *Plant Soil* **321**, 5–33 (2009).
22. V. V. S. R. Gupta, J. J. Germida, Soil aggregation: Influence on microbial biomass and implications for biological processes. *Soil Biol. Biochem.* **80**, A3–A9 (2015).
23. A. J. Daly, J. M. Baetens, B. De Baets, Ecological diversity: Measuring the unmeasurable. *Mathematics* **6**, 119-147 (2018).
24. L. Jost, Entropy and diversity. *Opinion* **2**, 363–375 (2006).
25. A. Chao, C. H. Chiu, L. Jost, Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annu. Rev. Ecol. Evol. Syst.* **45**, 297–324 (2014).
26. E. K. Morris *et al.*, Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecol. Evol.* **4**, 3514–3524 (2014).
27. A. Mentges, T. Dittmar, B. Blasius, C. Feenders, M. Seibt, Functional molecular diversity of marine dissolved organic matter is reduced during degradation. *Front. Mar. Sci.* **4**, 1–10 (2017).
28. M. Seidel, S. Phani, B. Vemulapalli, D. Mathieu, T. Dittmar, Marine dissolved organic matter shares thousands of molecular formulae yet differs structurally across major water masses. *Environ. Sci. Technol.* **56**, 3758–3769 (2022).
29. A. M. Kellerman, T. Dittmar, D. N. Kothawala, L. J. Tranvik, Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 1–8 (2014).
30. V. N. Roth *et al.*, Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).

31. Y. Ding *et al.*, Chemodiversity of soil dissolved organic matter. *Environ. Sci. Technol.* (2020), 10.1021/acs.est.0c01136.
32. E. Defossez *et al.*, Spatial and evolutionary predictability of phytochemical diversity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, 1–7 (2021).
33. D. C. Coleman, W. B. Whitman, Linking species richness, biodiversity and ecosystem function in soil systems. *Pedobiologia (Jena).* **49**, 479–497 (2005).
34. Z. Botta-Dukat, Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *J. Veg. Sci.* **16**, 533–540 (2005).
35. R. Rao, Diversity and Dissimilarity. *Theor. Popul. Biol.* **21**, 24–43 (1982).
36. O. L. Petchey, K. J. Gaston, Functional diversity: Back to basics and looking forward. *Ecol. Lett.* **9**, 741–758 (2006).
37. K. Kalbitz *et al.*, Changes in properties of soil-derived dissolved organic matter induced by biodegradation. *Soil Biol. Biochem.* **35**, 1129–1142 (2003).
38. J. Sanderman, J. A. Baldock, R. Amundson, Dissolved organic carbon chemistry and dynamics in contrasting forest and grassland soils. *Biogeochemistry* **89**, 181–198 (2008).
39. W. Amelung, S. Brodowski, A. Sandhage-Hofmann, R. Bol, *Chapter 6 Combining Biomarker with Stable Isotope Analyses for Assessing the Transformation and Turnover of Soil Organic Matter* (Elsevier Inc., ed. 1, 2009).
40. C. Kramer, G. Gleixner, Soil organic matter in soil depth profiles: Distinct carbon preferences of microbial groups during carbon transformation. *Soil Biol. Biochem.* **40**, 425–433 (2008).
41. M. Keiluweit, T. Wanzek, M. Kleber, P. Nico, S. Fendorf, Anaerobic microsites have an unaccounted role in soil carbon stabilization. *Nat. Commun.* **8**, 1–8 (2017).
42. M. T. Madigan, J. M. Martinko, K. S. Bender, D. H. Buckley, D. A. Stahl, *Brock Biology of Microorganisms* (Pearson Education Inc, ed. 14, 2015).
43. I. Kögel-Knabner, The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter: Fourteen years on. *Soil Biol. Biochem.* **34**, 139–162 (2002).
44. O. Pisani *et al.*, Accumulation of aliphatic compounds in soil with increasing mean annual temperature. *Org. Geochem.* **76**, 118–127 (2014).
45. I. Kögel-Knabner, The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter: Fourteen years on. *Soil Biol. Biochem.* **105**, A3–A8 (2017).
46. A. J. Simpson, M. J. Simpson, E. Smith, B. P. Kelleher, Microbially derived inputs to soil organic matter: Are current estimates too low? *Environ Sci. Technol.* **42**, 3115 (2008).
47. A. Miltner, P. Bombach, B. Schmidt-Brücken, M. Kästner, SOM genesis: Microbial biomass as a significant source. *Biogeochemistry* **111**, 41–55 (2012).
48. C. Rumpel, I. Kögel-Knabner, F. Bruhn, Vertical distribution, age, and chemical composition of organic carbon in two forest soils of different pedogenesis. *Org. Geochem.* **33**, 1131–1142 (2002).
49. M. G. Kramer *et al.*, The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls. *Annu. Rev. Ecol. Evol. Syst.* **48**, 419–445 (2017).
50. X. Zhu, R. D. Jackson, E. H. DeLucia, J. M. Tiedje, C. Liang, The soil microbial carbon pump: From conceptual insights to empirical assessments. *Glob. Chang. Biol.* **26**, 6032–6039 (2020).
51. J. Lehmann, M. Kleber, The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
52. T. Camenzind, K. Mason-Jones, I. Mansour, M. C. Rillig, J. Lehmann, Formation of necromass-derived soil organic carbon determined by microbial death pathways. *Nat. Geosci.*, (2023), 10.1038/s41561-022-01100-3.
53. G. Gleixner, Soil organic matter dynamics: A biological perspective derived from the use of compound-specific isotopes studies. *Ecol. Res.* **28**, 683–695 (2013).
54. H. Kang, W. Yu, S. Dutta, H. Gao, Soil microbial community composition and function are closely associated with soil organic matter chemistry along a latitudinal gradient. *Geoderma* **383**, 114744 (2021).
55. H. Y. Li *et al.*, The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. *Microbiome* **6**, 1–16 (2018).
56. C. A. Creamer *et al.*, Mineralogy dictates the initial mechanism of microbial necromass association. *Geochim. Cosmochim. Acta* **260**, 161–176 (2019).
57. J. Lehmann *et al.*, Spatial complexity of soil organic matter forms at nanometre scales. *Nat. Geosci.* **1**, 238–242 (2008).
58. E. K. Coward, T. Ohno, A. F. Plante, Adsorption and molecular fractionation of dissolved organic matter on iron-bearing mineral matrices of varying crystallinity. *Environ. Sci. Technol.* **52**, 1036–1044 (2018).
59. N. Eisenhauer *et al.*, Root biomass and exudates link plant diversity with soil bacterial and fungal biomass. *Sci. Rep.* **7**, 1–8 (2017).
60. M. Lange *et al.*, Plant diversity increases soil microbial activity and soil carbon storage. *Nat. Commun.* **6**, 1–8 (2015).
61. N. Fierer, J. P. Schimel, P. A. Holden, Variations in microbial community composition through two soil depth profiles. *Soil. Biol. Biochem.* **35**, 167–176 (2003).
62. K. Heckman, C. R. Lawrence, J. W. Harden, A sequential selective dissolution method to quantify storage and stability of organic carbon associated with Al and Fe hydroxide phases. *Geoderma* **312**, 24–35 (2018).
63. S. E. Fick, R. J. Hijmans, WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
64. T. Regier *et al.*, Performance and capabilities of the Canadian Dragon: The SGM beamline at the Canadian Light Source. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **582**, 93–95 (2007).
65. B. Ravel, M. Newville, ATHENA, ARTEMIS, HEPHAESTUS: Data analysis for X-ray absorption spectroscopy using IFEFFIT. *J. Synchrotron Radiat.* **12**, 537–541 (2005).
66. M. Schumacher *et al.*, Chemical composition of aquatic dissolved organic matter in five boreal forest catchments sampled in spring and fall seasons. *Biogeochemistry* **80**, 263–275 (2006).
67. J. Lehmann *et al.*, Near-edge X-ray absorption fine structure (NEXAFS) spectroscopy for mapping nano-scale distribution of organic carbon forms in soil: Application to black carbon particles. *Global Biogeochem. Cycles* **19**, 1–12 (2005).
68. M. Wang *et al.*, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
69. L. F. Nothias *et al.*, Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* **17**, 905–908 (2020).
70. T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
71. M. Katajamaa, J. Miettinen, M. Orešič, MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–636 (2006).
72. A. Chao *et al.*, Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters. *Ecol. Monogr.* **84**, 45–67 (2014).
73. H. Lenth *et al.*, emmeans: Estimated Marginal Means, aka Least-Squares Means. *R Stud.* (2022).
74. M. J. Anderson, D. C. I. Walsh, PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecol. Monogr.* **83**, 557–574 (2013).