Supporting Information for

Decomposition Decreases Molecular Diversity and Ecosystem Similarity of Soil Organic Matter

Rachelle Davenport[1], Benjamin P. Bowen[2,3], Laurel M. Lynch[1], Suzanne M. Kosina[2], Itamar Shabtai[1], Trent R. Northen[2,3], and Johannes Lehmann[1,4,5,6]*

[1]Soil and Crop Sciences, School of Integrative Plant Sciences, Cornell University, 306 Tower Rd, Ithaca, NY, 14850, USA.
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA, 94720, USA.
[3]DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA, 94598, USA.
[4]Department of Global Development, Cornell University, Ithaca, NY, 14850, USA.
[5]Cornell Institute for Digital Agriculture CIDA, Cornell University, Ithaca, NY, 14850, USA.
[6]Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY, 14850, USA.
*Corresponding author
Email: CL273@cornell.edu

This PDF file includes: Supporting text; Figures S1 to S9; Tables S1 to S10; SI References

Supporting Information Text

Ecological diversity indices for mass spectrometry data

The importance of biodiversity for ecosystem health is well established (1–4), as are mathematical functions describing biodiversity at the macro- and micro-scale. Indices quantifying molecular diversity in aquatic and ecological studies have been developed for decades (5–8), but the exploration of molecular diversity of soil organic matter remains in its infancy. We reviewed numerous diversity indices (9–13) to identify the most appropriate metrics for molecular data measured using mass spectrometry that are consistent with expectations for what aspects of diversity is relevant with respect to both biotic production and microbial consumption of organic compounds. Here, we outline the commonly used diversity indices and their usefulness in describing molecular diversity.

(1) Molecular richness, defined as the absolute number of identified molecules;

$$D_R = S \qquad \text{eq. 1.}$$

where S is the total number of molecules in the community and the classification of each molecule is known. Molecular richness considers only presence/absence data, is influenced by sampling size, and does not consider the relative abundance, evenness, or dissimilarity among molecules.

(2) Shannon's Diversity (H_{Sh}) provides a measure of entropy, calculated as,

$$D_{Sh} = -\sum_{i=1}^{S} p_i \ln p_i \qquad \text{eq. 2.}$$

where S is the total number of molecules in the community, and p_i is the relative abundance of molecules (9, 14). While Shannon's diversity accounts for richness and evenness, it gives more weight to common species and does not assess molecular dissimilarity. As an entropy equation, Shannon's Diversity estimates the inherent uncertainty in predicting the identity of a given molecule in a complex mixture.

(3) The Simpson Diversity (H_{Si}) index is a probability equation, calculated as,

$$D_{Si} = \sum_{i=1}^{S} p_i^2 \qquad \text{eq. 3.}$$

where S equals the total number of molecules in the community and pi is the relative abundance of molecules (9, 14). The Simpson Diversity index represents the probability that two randomly selected molecules from a mixture represent the same species. While the Simpson index does account for richness and evenness, it does not include molecular dissimilatory information. The Simpson Diversity index is often reported as the Gina-Simpson, (D_{GS} = 1-D_{Si}) or the Simpson Dominance Index (D_{SD} = 1/D_{Si}).

(4) The Chao 1 equation has been used as a non-parametric estimate of molecular richness, and is calculated as,

$$Chao1 = S_{obs} + \left(\frac{a^2}{2b}\right) \qquad \text{eq. 4}$$

where S_{obs} is the observed number of molecules, a is the number of observed molecules represented by a single molecule in the sample (singletons) and b is the number of observed molecules that occur twice in the sample (doubletons) (15, 16). The Chao 1 index does not include molecular evenness or molecular dissimilatory information. Furthermore, when analyzing mass spectrometry data and converting peak heights of identified features to relative abundances there is no certain way to determine which molecules occur once or twice in a sample.

(5) Hill Numbers represent a unifying diversity index that incorporates molecular richness and evenness and is calculated as,

$$ {}^{q}D_{H}(p) = \left(\sum_{i=1}^{S} p_i^{q}\right)^{1/(1-q)} \qquad \text{eq. 5.} $$

Where p_i is the relative abundance of molecules and $q[0,\infty]$ is the order of diversity, which indicates the sensitivity to rare or common molecules (11, 12, 14, 17, 18). When q=0 ($D_{H,q=0}$), there is no sensitivity to the relative abundance of molecules, thus equals molecular richness. When q=1 ($D_{H,q=1}$),, all molecules are equally weighted by frequency without favoring any, and when q=2 ($D_{HN,q=2}$),, common species are favored. Hill numbers have several strengths when applied to the molecular diversity of DOM: the equation (1) the effective number of molecules, or estimated number of molecules per sample, and (2) values can be compared across studies (unlike most classical diversity indices).

(6) Rao's quadratic entropy is a similarity-sensitive diversity index, defined as,

$$ FD_{Rao} = \sum_{i,j=1}^{S} d_{ij}\, p_i p_j \qquad \text{eq. 6.} $$

Where, d_{ij} is the dissimilarity between molecule i and j (or the inter-species diversity) and p_i is the relative abundance of molecule i and p_j is the relative abundance of molecule j (10, 19). Rao's quadratic entropy serves as a "functional" diversity index that considers the molecular similarity, based on an identifiable and ecological relevant molecular property. Two properties that are commonly assessed include molecular weight and the nominal oxidation state of carbon (NOSC) (8). FD_{Rao} also considered the relative abundances of molecules. As a result, FD_{Rao} values estimate the expected dissimilarity between two randomly selected molecules within a sample.

The previous six equations are highlighted since they are the most commonly used diversity indices in aquatic scientific studies (7, 8, 20, 21), with the exception of Hill Numbers. However, the use of such common indices may not be meaningful for molecules identified with mass spectrometry. For example, determining which compounds occur exactly once or twice as in the Chao 1 index would be highly uncertain and rely upon arbitrary conversions of peak heights to number of occurrences. Additionally, molecular richness is difficult to quantify given the detection limitations of mass spectrometry instruments or for that matter any analytical tools quantifying molecular properties in natural soil organic matter, and even more difficult to positively identify individual molecules. From the theoretical considerations above, we expect that the most appropriate diversity indices for molecular diversity that is relevant for soil organic matter studies are the Hill Numbers and Rao's quadratic entropy. Hill Numbers have the ability to be transformed into traditional Shannon and Simpson diversity indices, and obey the replication principle, meaning pooled assemblages have a linear property of diversity (11, 22). Using Hill Numbers, when q = 0, 1, and 2, enables us to investigate (i) molecular richness (favors common molecules), (ii) the diversity of evenly weighted molecules, and (iii) the diversity of rare molecules.

Rao's quadratic entropy further enables us to understand how the functional diversity of molecules is impacted by ecosystem properties. Functional diversity indices have become more commonly used by ecologists in an attempt to better understand ecosystems based on what organisms do, not just how many organisms are present (23). In the context of soil organic matter, this concept can be applied to not only which molecules are present, but how they interact with minerals or microbes in the soil. Thus, using functional diversity indices for molecules relies upon the selection of a molecular property that is ecologically relevant, which can vary based on the aim of specific research questions.

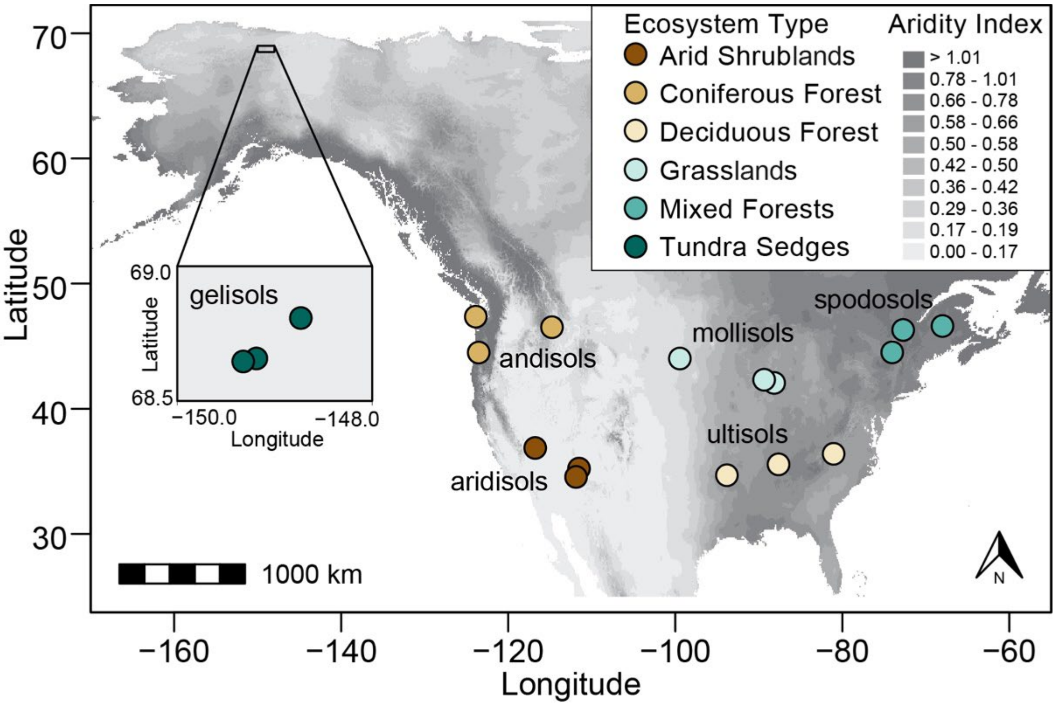

Figure S1. Map of selected soil profiles and the relationship between soil organic carbon and soil moisture. (A) Locations (n=18) were grouped into representative ecosystem types (n=6), based on dominant vegetation classes, and soil order, encompassing arid shrubs in the Southwest, coniferous forests in the Pacific Northwest, deciduous forests in the Southeast, grasses in the Midwest, mixed coniferous and hardwood forests in the Northeast, and tundra sedges in the Alaskan arctic.

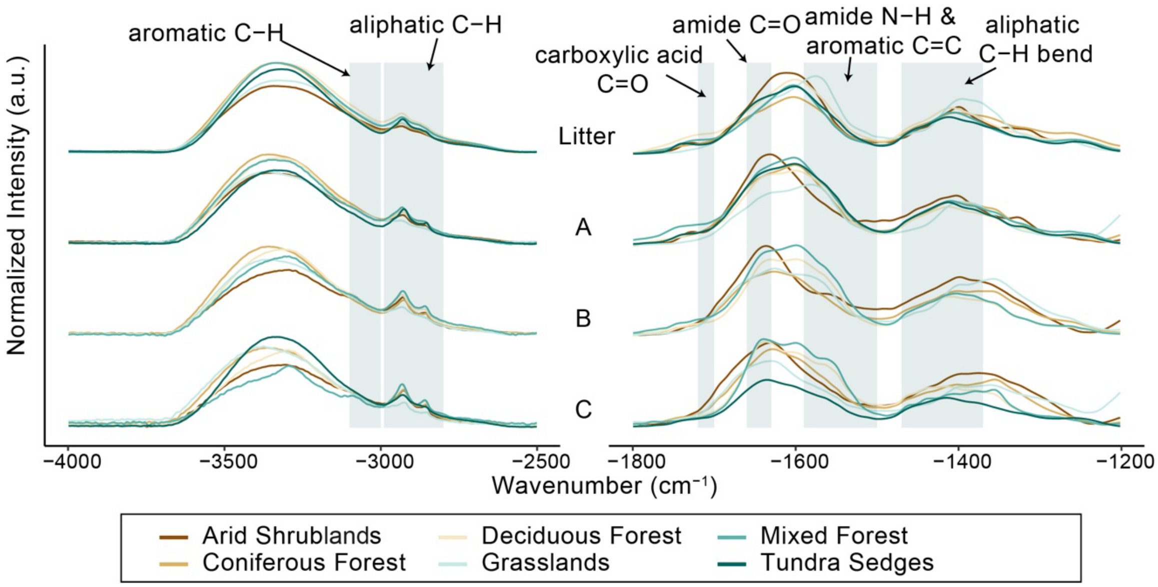

Figure S2. Fourier transform infrared attenuated total reflection (FTIR-ATR) spectra of water extractable organic matter from six ecosystem types. Spectra are means of three replicate locations. Shaded boxes indicate regions of interest, including aromatic C-H bonds (-3100 to -3000 cm^{-1}), aliphatic C-H bonds (-2990 to -2800 cm^{-1}), carboxylic acid C=O bonds (-1720 to -1700 cm^{-1}), amide C=O bonds (-1660 to -1630 cm^{-1}), amide N-H and aromatic C=C bonds (-1590 to -1500 cm^{-1}) and the aliphatic C-H bend region (-1470 to -1370 cm^{-1}).

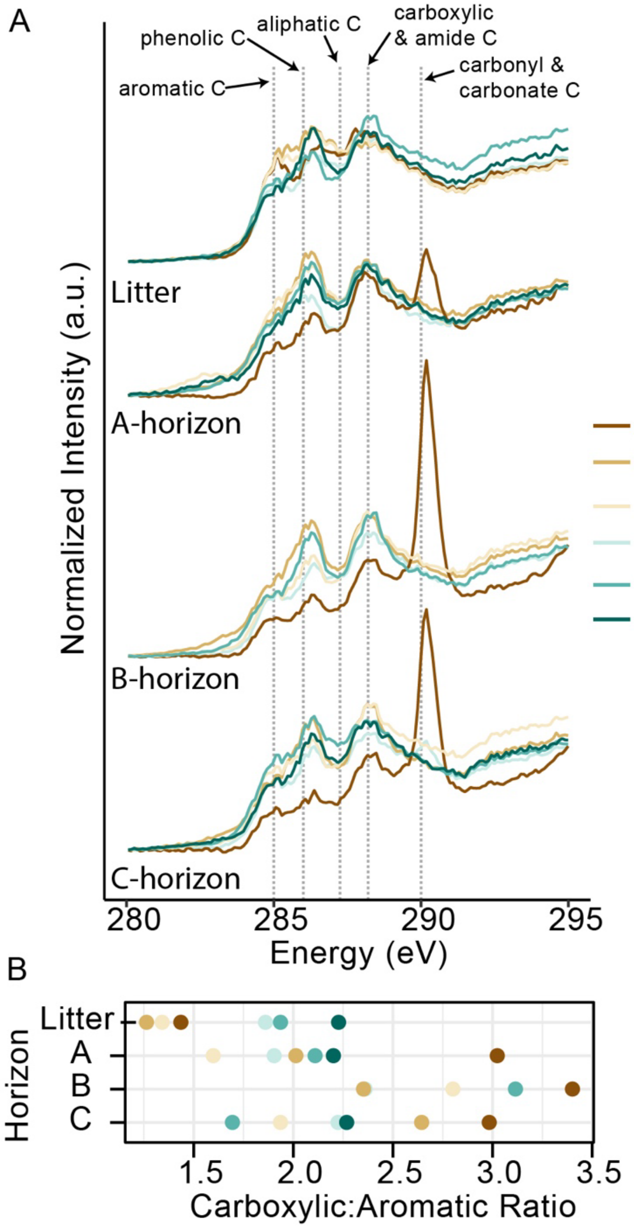

Figure S3. Near edge X-ray absorption fine-structure (NEXAFS) spectra for selected soil profiles from six ecosystem types. Spectra are from an arid shrubland, coniferous forest, deciduous forest, grassland, mixed coniferous and deciduous forest, and tundra tussock and sedges (A). Spectra were collected from the top litter layer, A, B, and C horizons. Carboxylic (288.2 eV) to aromatic (285.0 eV) ratios for each spectra indicate the degree of oxidation, with larger values suggesting SOM that is more oxidized. Dotted lines were drawn at the following energy levels: aromatic C 285.0 eV, phenolic C 286.0 eV, aliphatic C 287.3 eV, carboxylic and amide C 288.2 eV, and carbonyl and carbonate C 290.0 eV.

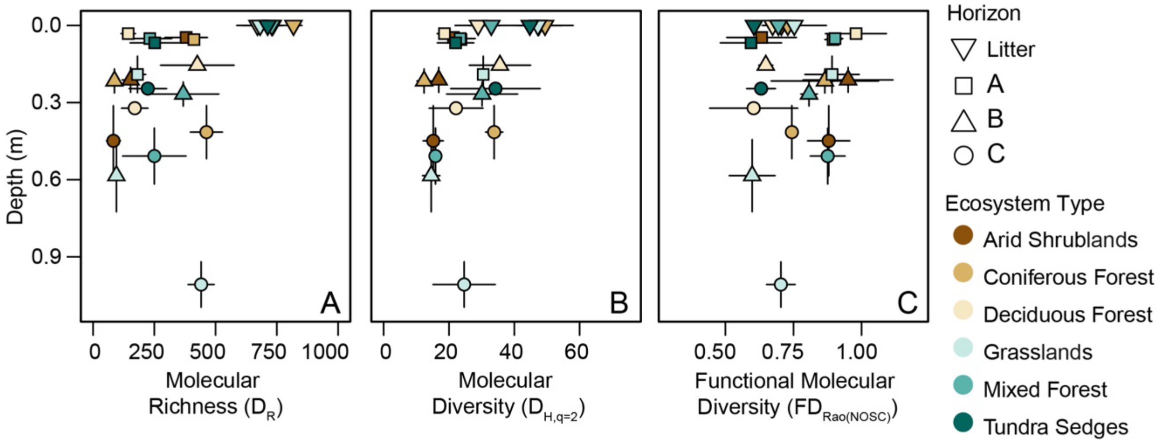

Figure S4. Molecular diversity of hydrophilic compounds throughout soil profiles under six ecosystem types. Shown are (A) molecular richness (D_R), (B) abundance-based molecular diversity ($D_{H,q=2}$), (C) functional molecular diversity using NOSC as molecular property of dissimilarity ($FD_{Rao(NOSC)}$).

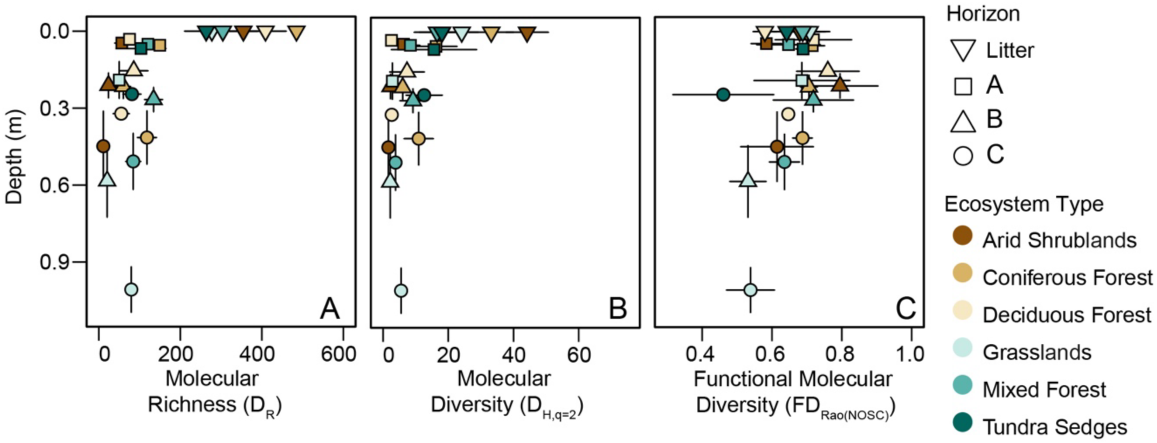

Figure S5. Molecular diversity of hydrophobic compounds throughout soil profiles under six ecosystem types. Shown are (A) molecular richness (D_R), (B) abundance-based molecular diversity ($D_{H,q=2}$), (C) functional molecular diversity using NOSC as molecular property of dissimilarity ($FD_{Rao(NOSC)}$).

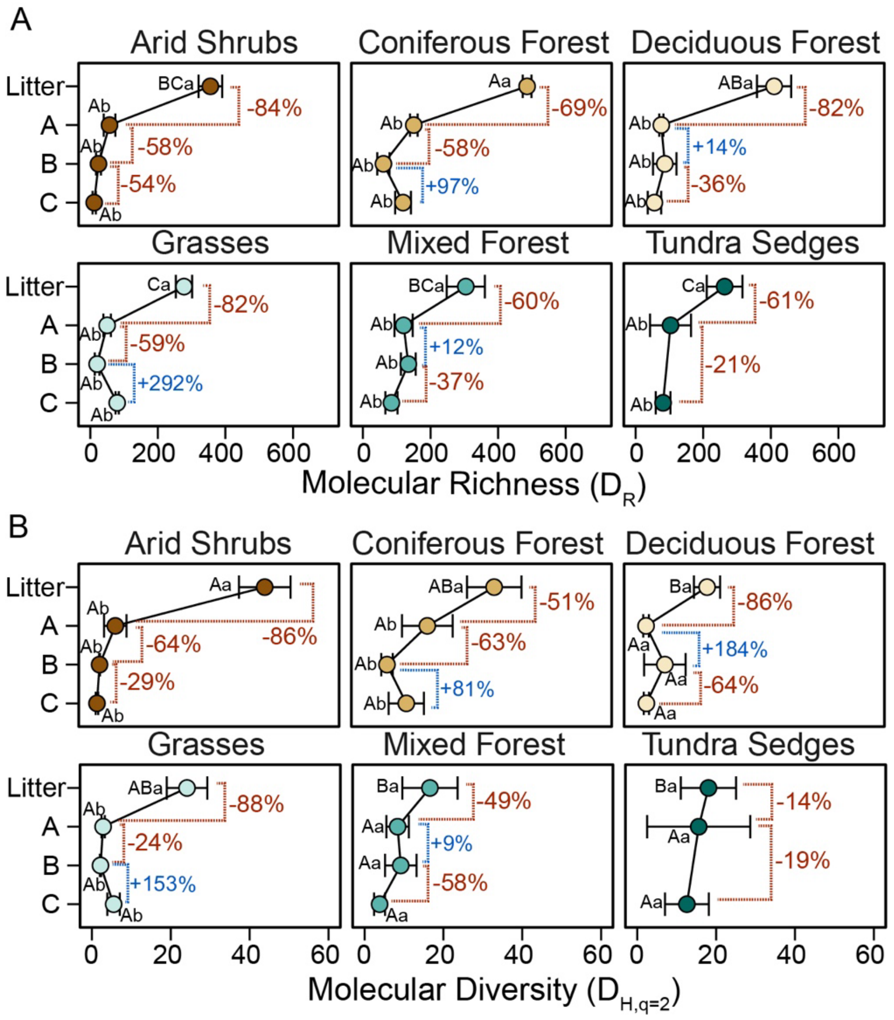

Figure S6. Molecular richness (A) and molecular diversity (B) of the litter, A, B, and C horizons of hydrophobic compounds within six ecosystem types. Percent differences between the litter and A-horizons, A- and B-horizons, and B- and C-horizons show the percent increase (blue) or decrease (red) in 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 vegetation classes. Vegetation classes that do not share the same capital letter are significantly different. Lower case 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 respectively. Trends for the hydrophilic compounds were similar, but stronger, than those of the hydrophobic compounds (Fig. 4).

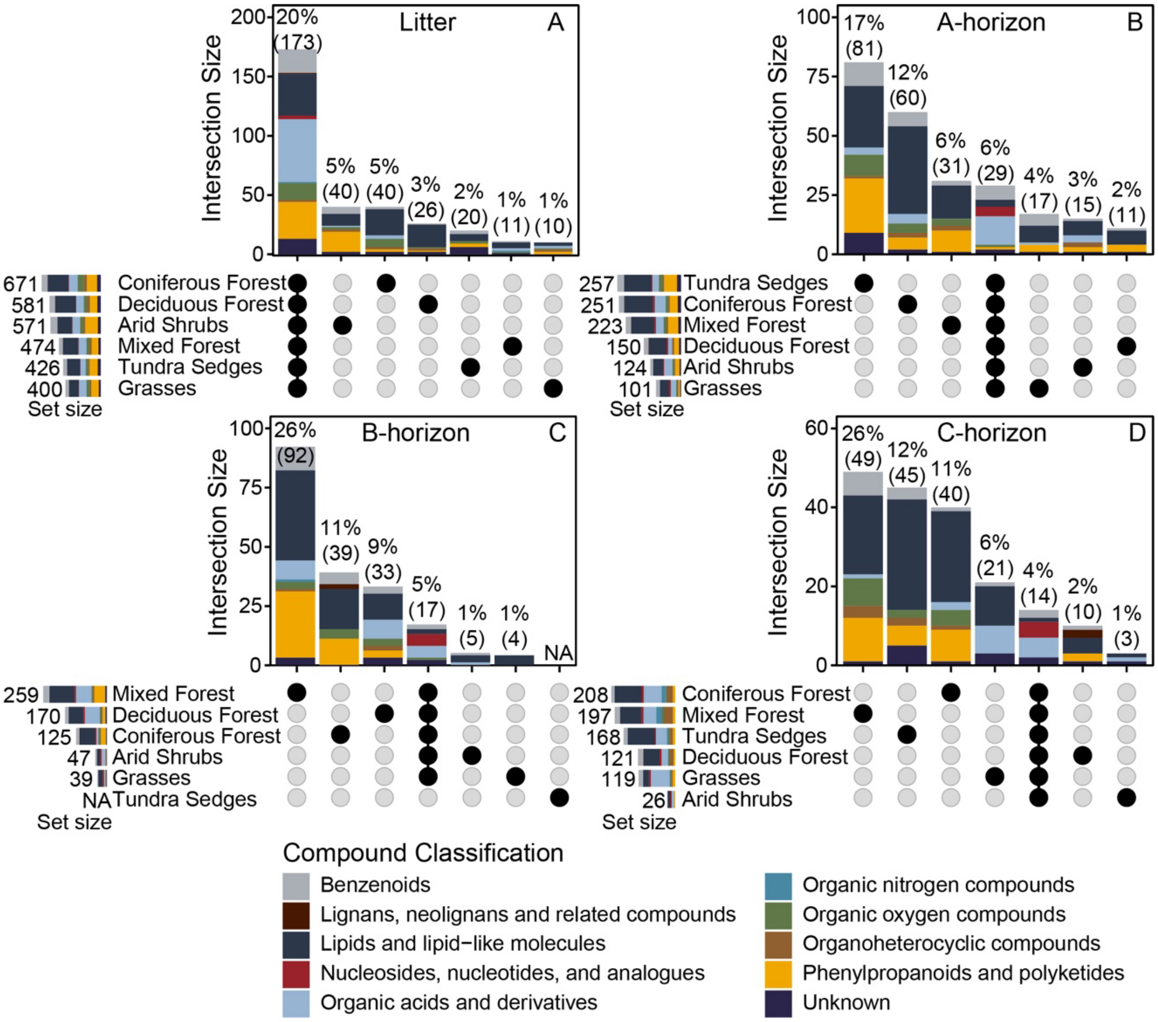

Figure S7. Shared and unique hydrophobic compounds identified in each horizon. Both shared and unique compounds are displayed for the (A) litter, (B) A-horizon, (C) B-horizon, (D) C-horizon. Black dots under vertical bars indicate sets of ecosystems considered; either as individual ecosystems (single 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) are shown. The proportion of compounds that are either shared or unique are 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 compounds were classified into superclass groupings and 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 hydrophilic compounds showed similar trends and are displayed in Fig. 1.

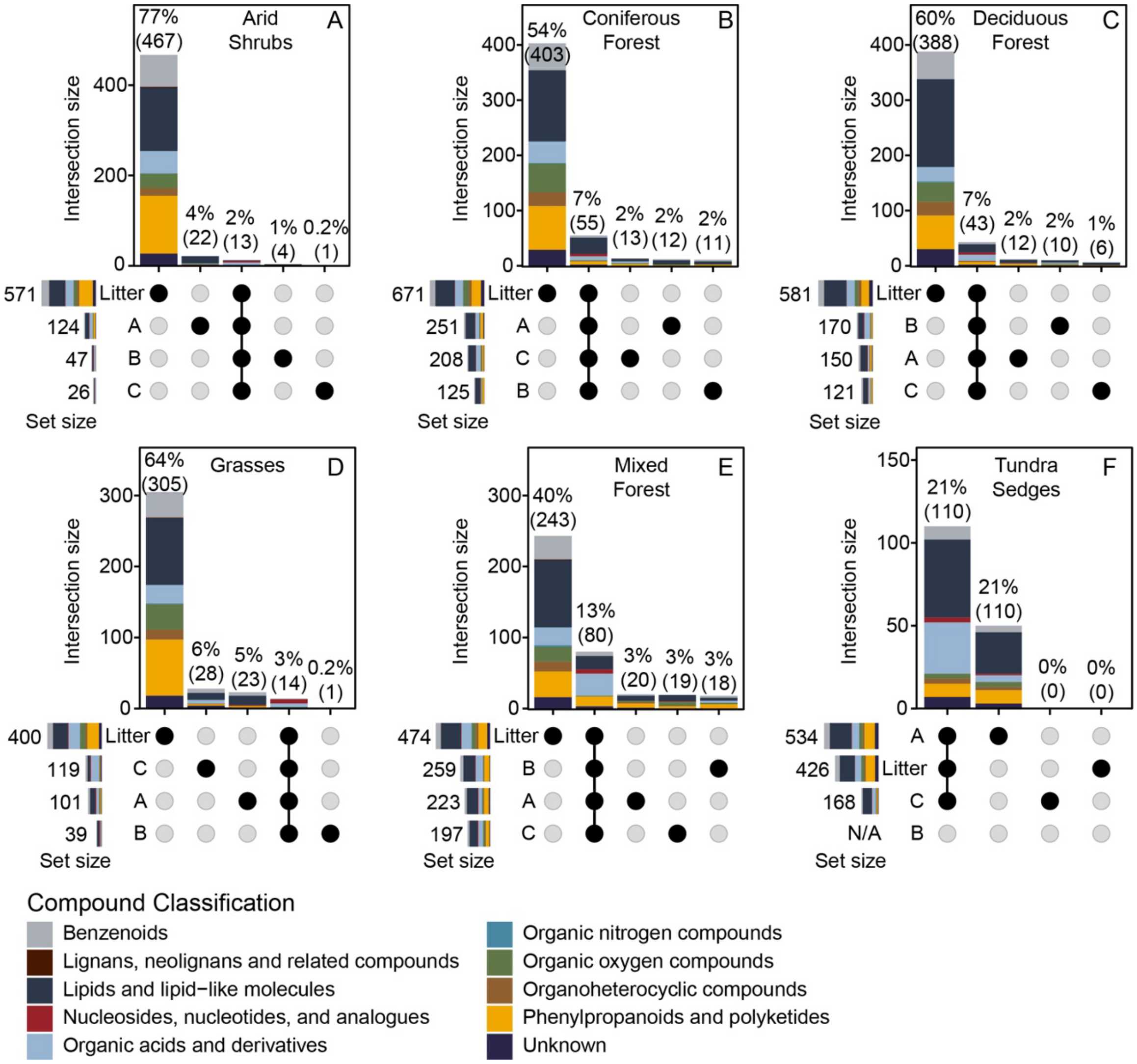

Figure S8. Shared and unique hydrophobic compounds identified in each ecosystem type Both shared and unique compounds are displayed for the litter, A-horizon, B-horizon, and C-horizon for the six ecosystem types: (A) arid shrubland, (B) coniferous forest, (C) deciduous forest, (D) grassland, (E) mixed 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) are shown. The proportion of compounds that are either shared or unique are 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 compounds were classified into superclass groupings and reported by color within both the vertical and horizontal bars. Proportion of features common or unique of the sum of features in each ecosystem type are displayed (proportions missing to 100% are features that are neither common nor unique; NA not available). Shared and unique features for hydrophilic compounds showed similar trends and are displayed in Fig. 2.

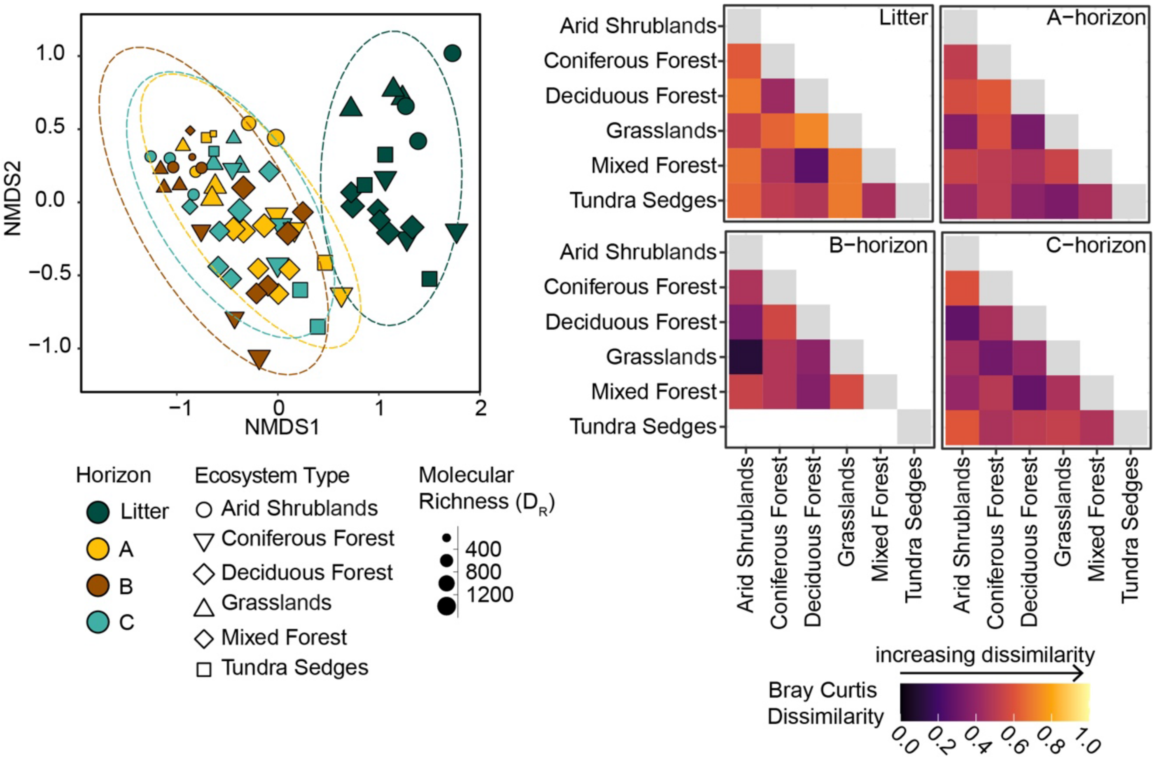

Figure S9. Distribution of hydrophobic compounds using non-metric dimensional scaling (NMDS) ordination (stress = 0.14) (A) and Bray Curtis Dissimilarity matrices (B). Potential predictor variables included as vectors in the NMDS ordination were mean annual temperature (MAT, °C), mean annual precipitation (MAP, mm), latitude, longitude, elevation (m), depth (m), concentrations of soil organic carbon (SOC), total nitrogen (TN), clay content, hydroxylamine extractable iron (mg g^{-1} soil), pH, gravimetric moisture content (GWC). No vectors were found to significantly explain ordination variance (p-value < 0.05). PERMANOVA analysis show that ecosystem type explained 17% ($p < 0.001$) and horizon explained 17% ($p < 0.001$) of the variance in dissimilarity of LCMS/MS SOM samples. The interaction of ecosystem type and horizon explained 20% 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 from the HILIC column, however the litter showed less dissimilarity compared to the C18 column and can be found in the main text (Fig. 5).

Table S1. Ecosystem designation of six regions.

Ecosystem Type	Soil Order	Region	Number of Sampling Sites	Mean Aridity	Mean Elevation (m)
Arid Shrublands	Aridisol	Southwest	3	0.12 ± 0.05	1287 ± 314
Coniferous Forest	Andisol	Pacific Northwest	3	1.69 ± 0.80	717 ± 244
Deciduous Forest	Ultisol	Southeast	3	0.75 ± 0.03	363 ± 87
Grasslands	Mollisol	Midwest	3	0.49 ± 0.12	326 ± 82
Mixed Forest	Spodosol	Northeast	3	0.87 ± 0.04	244 ± 152
Tundra Sedges	Gelisol	Alaskan Tundra	3	0.37 ± 0.01	677 ± 118

Table S2. Soil biogeochemistry properties of six ecosystem biomes by ecosystem type. Reported values are means ± standard errors of three representative soil profiles for each ecosystem.

Ecosystem	**Horizon**	**pH**	**Silt (%)**	**Clay (%)**	**Fe* (mg g^{-1} soil)**	**Al* (mg g^{-1} soil)**	**OC (mg g^{-1} litter or soil)**	**TN (mg g^{-1} litter or soil)**	**C:N Ratio**	**DOC* (mg g^{-1} soil)**
Arid Shrubs	Litter	NA	NA	NA	NA	NA	47.2 ± 2.2^{Aa}	1.3 ± 0.2^{Aa}	37.1 ± 6.0^{ABa}	NA
	A	6.84 ± 0.3^{Ab}	27 ± 4.1^{Aa}	18 ± 8.1^{Aa}	34.6 ± 10.2^{ABa}	34.3 ± 3.8^{Aa}	1.8 ± 0.8^{Ba}	0.2 ± 0.06^{Ba}	8.8 ± 1.8^{Ba}	1.3 ± 0.2^{Ca}
	B	7.83 ± 0.2^{Aa}	21 ± 4.7^{Aa}	11 ± 3.2^{Aa}	42.6 ± 17.5^{ABa}	38.6 ± 4.6^{ABa}	0.7 ± 0.3^{Ba}	0.0 ± 0.01^{Bb}	14.6 (± 3.1^{Aa}	0.7 ± 0.2^{Ca}
	C	7.25 ± 0.3^{Aab}	22 ± 6.3^{Aa}	17 ± 6.8^{ABa}	33.3 ± 19.3^{ABa}	30.9 ± 9.9^{ABa}	0.7 ± 0.3^{Ba}	0.1 ± 0.02^{Bb}	10.9 ± 4.3^{Aa}	0.7 ± 0.2^{Ca}
Coniferous Forest	Litter	NA	NA	NA	NA	NA	52.3 ± 2.0^{Aa}	0.7 ± 0.1^{Aa}	71.8 ± 9.3^{Aa}	NA
	A	4.39 ± 0.4^{BCa}	25 ± 4.8^{Aa}	13 ± 3.8^{Aa}	129.1 ± 46.1^{Aa}	117.2 ± 50.6^{Aa}	12.5 ± 2.5^{Aa}	0.5 ± 0.09^{ABa}	23.4 0.8^{ABa}	13.1 ± 4.7^{Aa}
	B	4.78 ± 0.3^{Ba}	22 ± 1.1^{Aa}	19 ± 6.5^{Aa}	161.0 ± 56.8^{Aa}	142.7 ± 59.1^{Aa}	6.3 ± 2.2^{Aab}	0.3 ± 0.08^{Aab}	23.3 ± 2.1^{Aa}	8.1 ± 2.1^{Aa}
	C	4.52 ± 0.1^{Ba}	25 ± 2.7^{Aa}	16 ± 5.7^{ABa}	134.1 ± 42.3^{Aa}	137.3 ± 58.8^{Aa}	3.6 ± 1.9^{ABb}	0.2 ± 0.09^{Bb}	18.9 ± 1.2^{Aa}	6.1 ± 1.6^{ABa}
Deciduous Forest	Litter	NA	NA	NA	NA	NA	50.8 ± 1.1^{Aa}	1.4 ± 0.2^{Aa}	38.7 ± 4.9^{ABa}	NA
	A	3.92 ± 0.3^{Ca}	48 ± 12.4^{Aa}	13 ± 1.5^{Aa}	16.3 ± 7.3^{Ba}	21.1 ± 2.9^{Aa}	5.0 ± 0.6^{ABa}	0.3 ± 0.00^{ABa}	18.7 ± 2.2^{ABa}	2.9 ± 0.3^{BCa}
	B	4.30 ± 0.3^{Ba}	38 ± 9.1^{Ab}	14 ± 3.0^{Aa}	13.9 ± 6.1^{Ba}	20.6 ± 4.1^{Ba}	1.2 ± 0.1^{ABb}	0.1 ± 0.00^{ABb}	13.2 ± 0.4^{Aa}	1.5 ± 0.3^{BCa}
	C	4.13 ± 0.1^{Ba}	41 ± 6.4^{Aab}	18 ± 2.9^{ABa}	56.0 ± 30.2^{Ba}	25.5 ± 6.1^{Ba}	0.6 ± 0.04^{Bb}	0.1 ± 0.01^{Bb}	10.6 ± 0.7^{Aa}	1.4 ± 0.2^{Ca}

Grasses	Litter	NA	NA	NA	NA	NA	47.6 ± 1.2^{Aa}	1.8 ± 0.1^{Aa}	26.4 ± 3.0^{Ba}	NA
	A	5.93 ± 0.5^{ABb}	43 ± 8.5^{Aa}	30 ± 6.0^{Aa}	36.6 ± 0.6^{Aba}	21.3 ± 4.2^{Aab}	3.3 ± 0.5^{ABa}	0.3 ± 0.03^{ABa}	9.7 ± 0.5^{Bab}	1.9 ± 0.5^{Cab}
	B	6.77 ± 0.4^{Aa}	39 ± 6.5^{Aa}	37 ± 8.8^{Aa}	60.0 ± 13.8^{ABa}	32.1 ± 7.0^{ABa}	1.1 ± 0.2^{ABa}	0.1 ± 0.02^{ABb}	8.9 ± 0.2^{Ab}	1.0 ± 0.3^{BCb}
	C	6.84 ± 0.5^{Aa}	42 ± 9.5^{Aa}	40 ± 10.9^{Aa}	44.7 ± 10.6^{ABa}	19.8 ± 3.8^{Bb}	1.7 ± 0.8^{Ba}	0.1 ± 0.01^{Bb}	27.2 ± 12.9^{Aa}	0.7 ± 0.2^{Ca}
Mixed Forest	Litter	NA	NA	NA	NA	NA	53.4 ± 2.4^{Aa}	1.1 ± 0.1^{Aa}	47.9 ± 6.6^{ABa}	NA
	A	4.04 ± 0.2^{Cb}	29 ± 7.1^{Aa}	8 ± 2.3^{Aa}	51.4 ± 11.0^{ABa}	34.3 ± 3.7^{Ab}	4.9 ± 0.1^{ABa}	0.3 ± 0.01^{ABa}	16.4 ± 1.1^{ABa}	4.1 ± 0.7^{ABCa}
	B	4.47 ± 0.1^{Bab}	23 ± 7.6^{Aa}	10 ± 3.8^{Aa}	43.9 ± 5.6^{ABa}	38.6 ± 14.0^{ABa}	1.7 ± 0.4^{ABab}	0.1 ± 0.02^{ABb}	15.6 ± 3.5^{Aa}	4.3 ± 1.5^{ABa}
	C	4.90 ± 0.4^{Ba}	25 ± 6.1^{Aa}	10 ± 3.2^{Ba}	37.5 ± 8.3^{ABa}	30.9 ± 7.1^{ABab}	0.7 ± 0.4^{Bb}	0.1 ± 0.2^{Bb}	12.3 ± 3.3^{Aa}	2.8 ± 1.3^{BCa}
Tundra Sedges	Litter	NA	NA	NA	NA	NA	31.0 ± 3.7^{Aa}	0.9 ± 0.1^{Aa}	36.1 ± 2.1^{ABa}	NA
	A	5.04 ± 0.7^{BCa}	NA	NA	78.5 ± 16.2^{ABa}	26.0 ± 2.7^{Aa}	23.9 ± 4.9^{A}	0.8 ± 0.20^{Aa}	34.5 ± 11.6^{Aa}	6.9 ± 2.8^{ABa}
	C	4.06 ± 0.2^{Bb}	NA	NA	121.7 ± 28.0^{Aa}	38.4 ± 1.9^{ABa}	21.6 ± 11.0^{A}	0.8 ± 0.39^{Aa}	23.3 ± 2.6^{Aa}	5.4 ± 1.5^{Aa}

Capital letters indicate differences across vegetation class for A, B or C horizon, lower case letters indicate differences within a single vegetation class with depth. Differences of means were determined with mixed effects models using a Bonferroni correction factor. NA not available.

*Iron, aluminum and DOC values reported are for hydroxylamine HCl extracts.

Table S3. Molecular diversity indices, NOSC, and molecular weight of identified hydrophilic compounds from LC-MS/MS HILIC column. Reported values are means ± standard errors.

Ecosystem	**Horizon**	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{H,q=1}$)**	**Molecular Diversity ($D_{H,q=2}$)**	**Molecular Diversity ($FD_{Rao(MW)}$)**	**Molecular Diversity ($FD_{Rao(NOSC)}$)**	**Molecular Weight (amu)**	**Nominal Oxidation State of Carbon (NOSC)**
Arid Shrub	Litter	733 ± 49^{Aa}	102 ± 8^{Aa}	47 ± 4^{Aa}	141.4 ± 3.9^{Aa}	0.71 ± 0.04^{Aa}	246.8 ± 3.1^{Ab}	-0.11 ± 0.01^{Aa}
	A	380 ± 87^{Ab}	56 ± 10^{Ab}	21 ± 3^{Ab}	175.7 ± 12.2^{Aa}	0.63 ± 0.13^{Aa}	327.3 ± 17.7^{Aab}	-0.30 ± 0.04^{Aa}
	B	153 ± 34^{ABbc}	35 ± 4^{Ab}	17 ± 1^{Ab}	167.6 ± 14.7^{Aa}	0.95 ± 0.17^{Aa}	349.0 ± 41.7^{Aa}	-0.30 ± 0.17^{Aa}
	C	83 ± 29^{Bc}	29 ± 7^{Bb}	15 ± 3^{Ab}	160.5 ± 23.2^{ABa}	0.88 ± 0.08^{Aa}	388.3 ± 24.5^{Aa}	-0.29 ± 0.16^{Aa}
Coniferous Forest	Litter	818 ± 35^{Aa}	117 ± 14^{Aa}	49 ± 9^{Aa}	117.6 ± 12.1^{Aa}	0.73 ± 0.04^{Aa}	249.4 ± 10.0^{Ab}	-0.11 ± 0.02^{Aa}
	A	412 ± 22^{Ab}	62 ± 3^{Abc}	24 ± 1^{Ab}	185.1 ± 17.4^{Aa}	0.90 ± 0.03^{Aa}	296.0 ± 1.0^{Aab}	-0.24 ± 0.02^{Aa}
	B	87 ± 6^{Ac}	24 ± 2^{Ac}	12 ± 2^{Ab}	113.9 ± 39.3^{Aa}	0.87 ± 0.20^{Aa}	355.0 ± 21.1^{Aa}	-0.04 ± 0.15^{Aa}
	C	463 ± 66^{ABab}	85 ± 2^{Aab}	34 ± 3^{Aab}	137.1 ± 24.9^{ABa}	0.74 ± 0.01^{Aa}	266.6 ± 17.9^{Bb}	-0.18 ± 0.05^{Aa}
Deciduous Forest	Litter	671 ± 62^{Aa}	79 ± 14^{Aa}	29 ± 7^{Aa}	83.2 ± 10.2^{Aa}	0.67 ± 0.02^{Aab}	227.9 ± 4.6^{Ab}	-0.07 ± 0.01^{Aa}
	A	143 ± 29^{Ac}	39 ± 2^{Aa}	19 ± 2^{Aa}	156.2 ± 8.2^{Aa}	0.98 ± 0.11^{Aa}	328.1 ± 1.2^{Aa}	-0.05 ± 0.13^{Aa}

	B	425 ± 150^{Aab}	76 ± 17^{Aa}	36 ± 9^{Aa}	113.9 ± 17.2^{Aa}	0.65 ± 0.03^{Aab}	270.6 ± 0.3^{Ab}	-0.11 ± 0.07^{Aa}
	C	170 ± 54^{ABbc}	45 ± 15^{ABa}	22 ± 8^{Aa}	137.1 ± 15.7^{ABa}	0.60 ± 0.16^{Ab}	310.4 ± 8.2^{ABab}	-0.12 ± 0.10^{Aa}
Grasses	Litter	682 ± 96^{Aa}	100 ± 12^{Aa}	47 ± 7^{Aa}	132.4 ± 21.4^{Aab}	0.75 ± 0.12^{Aa}	246.2 ± 10.8^{Aa}	-0.14 ± 0.02^{Aa}
	A	181 ± 33^{Abc}	59 ± 7^{Aab}	30 ± 2^{Aab}	172.9 ± 10.1^{Aa}	0.89 ± 0.10^{Aa}	332.2 ± 14.0^{Aa}	-0.10 ± 0.15^{Aa}
	B	95 ± 19^{Bab}	29 ± 4^{Aab}	15 ± 3^{Aab}	131.3 ± 13.4^{Ab}	0.60 ± 0.08^{Aa}	323.6 ± 22.4^{Aa}	-0.03 ± 0.12^{Aa}
	C	441 ± 53^{Ac}	70 ± 21^{ABb}	25 ± 10^{Ab}	76.9 ± 8.9^{Bab}	0.70 ± 0.05^{Aa}	249.8 ± 5.9^{Ba}	-0.09 ± 0.03^{Aa}
Mixed Forest	Litter	729 ± 36^{Aa}	88 ± 12^{Aa}	33 ± 7^{Aa}	85.0 ± 1.6^{Ac}	0.69 ± 0.02^{Aa}	230.8 ± 1.5^{Ab}	-0.09 ± 0.02^{Aa}
	A	231 ± 89^{Ab}	56 ± 6^{Aab}	23 ± 5^{Aa}	220.4 ± 26.7^{Aa}	0.90 ± 0.03^{Aa}	367.3 ± 24.1^{Aa}	-0.43 ± 0.10^{Aa}
	B	368 ± 145^{ABb}	77 ± 28^{Aab}	30 ± 11^{Aa}	141.7 ± 7.1^{Abc}	0.81 ± 0.03^{Aa}	332.2 ± 49.5^{Aa}	-0.27 ± 0.10^{Aa}
	C	250 ± 129^{ABb}	42 ± 12^{ABb}	16 ± 2^{Aa}	160.8 ± 37.5^{ABab}	0.88 ± 0.06^{Aa}	309.5 ± 31.7^{ABab}	-0.21 ± 0.13^{Aa}
Tundra Sedges	Litter	713 ± 14^{Aa}	113 ± 5^{Aa}	45 ± 3^{Aa}	155.1 ± 5.3^{Aa}	0.60 ± 0.02^{Aa}	250.0 ± 3.8^{Ab}	-0.06 ± 0.01^{Aa}
	A	252 ± 101^{ABb}	46 ± 1^{ABb}	22 ± 6^{Aa}	148.3 ± 32.2^{Aa}	0.59 ± 0.11^{Aa}	277.4 ± 27.1^{ABab}	-0.05 ± 0.12^{Aa}
	C	223 ± 77^{Ab}	68 ± 26^{Ab}	34 ± 14^{Aa}	187.1 ± 21.6^{Aa}	0.63 ± 0.05^{Aa}	341.7 ± 42.7^{Aa}	-0.23 ± 0.10^{Aa}

Capital letters indicate differences across ecosystems within a given horizon (Litter, A, B or C). Ecosystems that do not share the same capital letter can be considered significantly different. Lower case letters indicate differences between horizons within a single ecosystem. Horizons that do not share the same lowercase letters can be considered significantly different. Differences of means were determined with mixed effects models using a Bonferroni correction for 3 or 15 tests respectively.

Table S4. Molecular diversity indices, NOSC, and molecular weight of identified hydrophobic from LC-MS/MS C18 column. Reported values are means ± standard errors.

Ecosystem	**Horizon**	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{H,q=1}$)**	**Molecular Diversity ($D_{H,q=2}$)**	**Molecular Diversity ($FD_{Rao(MW)}$)**	**Molecular Diversity ($FD_{Rao(NOSC)}$)**	**Molecular Weight (amu)**	**Nominal Oxidation State of Carbon (NOSC)**
Arid Shrub	Litter	355 ± 34^{BCa}	97 ± 8^{Aa}	44 ± 7^{Aa}	153.9 ± 10.8^{Aa}	0.68 ± 0.08^{Aa}	280.0 ± 3.1^{Aa}	-0.50 ± 0.02^{Aa}
	A	56 ± 17^{Ab}	14 ± 8^{Ab}	6 ± 3^{Ab}	98.0 ± 12.6^{Aa}	0.58 ± 0.04^{Aa}	305.8 ± 6.9^{Aa}	-0.59 ± 0.06^{Aa}
	B	24 ± 7^{Ab}	4 ± 0.3^{Ab}	2 ± 0.1^{Ab}	106.4 ± 32.2^{Aa}	0.79 ± 0.11^{Aa}	284.6 ± 16.7^{ABa}	-0.47 ± 0.04^{Aa}
	C	11 ± 5^{Ab}	2 ± 1^{Ab}	2 ± 0.3^{Ab}	111.1 ± 29.3^{Aa}	0.61 ± 0.10^{Aa}	299.0 ± 18.1^{Aa}	-0.60 ± 0.09^{ABa}
Coniferous Forest	Litter	485 ± 13^{Aa}	97 ± 12^{Aa}	33 ± 7^{ABa}	112.8 ± 5.0^{Aa}	0.70 ± 0.04^{Aa}	260.0 ± 3.2^{Aa}	-0.62 ± 0.04^{Aa}
	A	150 ± 11^{Ab}	40 ± 10^{Ab}	16 ± 6^{Ab}	128.8 ± 27.8^{Aa}	0.72 ± 0.04	291.7 ± 3.7^{Aa}	-0.77 ± 0.04^{Aa}
	B	60 ± 18^{Ab}	11 ± 3^{Ab}	6 ± 1^{Ab}	89.2 ± 14.2^{Aa}	0.71 ± 0.02^{Aa}	296.6 ± 7.3^{Aa}	-0.82 ± 0.08^{Ba}
	C	118 ± 23^{Ab}	27 ± 9^{Ab}	11 ± 4^{Ab}	100.6 ± 12.6^{Aa}	0.69 ± 0.03^{Aa}	279.5 ± 14.2^{Aa}	-0.69 ± 0.07^{ABa}
Deciduous Forest	Litter	409 ± 50^{ABa}	78 ± 7^{Aa}	18 ± 3^{Ba}	99.3 ± 9.8^{Aab}	0.58 ± 0.03^{Aa}	256.7 ± 3.6^{Ab}	-0.72 ± 0.03^{Aa}
	A	75 ± 6^{Ab}	8 ± 3^{Ab}	3 ± 1^{Aa}	125.1 ± 4.7^{Aa}	0.72 ± 0.11^{Aa}	305.5 ± 9.5^{Aa}	-0.74 ± 0.06^{Aa}
	B	86 ± 35^{Ab}	17 ± 12^{Ab}	7 ± 5^{Aa}	64.1 ± 11.1^{Ab}	0.76 ± 0.09^{Aa}	268.3 ± 14.3^{ABab}	-0.58 ± 0.08^{ABa}
	C	55 ± 20^{Ab}	6 ± 2^{Ab}	3 ± 1^{Aa}	106.8 ± 8.7^{Aab}	0.65 ± 0.01^{Aa}	291.9 ± 13.9^{Aab}	-0.76 ± 0.02^{ABa}

Grasses	Litter	277 ± 24^{Ca}	56 ± 12^{Aa}	24 ± 5^{ABa}	132.5 ± 15.7^{Aa}	0.71 ± 0.06^{Aa}	281.2 ± 8.2^{Aa}	-0.48 ± 0.04^{Aa}
	A	50 ± 11^{Ab}	6 ± 1^{Ab}	3 ± 0.4^{Ab}	92.9 ± 11.8^{Aa}	0.69 ± 0.14^{Aa}	266.3 ± 19.0^{Aa}	-0.68 ± 0.06^{Aa}
	B	20 ± 7^{Ab}	3 ± 0.4^{Ab}	2 ± 0.2^{Ab}	83.0 ± 23.0^{Aa}	0.53 ± 0.05^{Aa}	238.3 ± 1.2^{Ba}	-0.65 ± 0.16A^{Ba}
	C	80 ± 5^{Ab}	12 ± 4^{Ab}	6 ± 2^{Ab}	82.4 ± 30.2^{Aa}	0.54 ± 0.07^{Aa}	251.2 ± 2.8^{Aa}	-0.51 ± 0.06^{ABa}
Mixed Forest	Litter	304 ± 56^{BCa}	60 ± 15^{Aa}	17 ± 7^{Ba}	102.4 ± 7.6^{Aa}	0.69 ± 0.06^{Aa}	253.5 ± 0.7^{Ab}	-0.64 ± 0.02^{Aa}
	A	120 ± 27^{Ab}	27 ± 9^{Ab}	8 ± 3^{Aa}	122.8 ± 4.3^{Aa}	0.65 ± 0.09^{Aa}	320.0 ± 8.4^{Aa}	-0.76 ± 0.09^{Aa}
	B	134 ± 22^{Ab}	27 ± 9^{Aab}	9 ± 4^{Aa}	116.6 ± 6.6^{Aa}	0.72 ± 0.11^{Aa}	311.5 ± 16.2^{Aa}	-0.74 ± 0.06^{ABa}
	C	84 ± 18^{Ab}	12 ± 5^{Ab}	4 ± 1^{Aa}	116.1 ± 17.7^{Aa}	0.64 ± 0.04^{Aa}	296.6 ± 22.8^{Aab}	-0.71 ± 0.13^{ABa}
Tundra Sedges	Litter	263 ± 53^{Ca}	56 ± 14^{Aa}	18 ± 7^{Ba}	149.3 ± 0.1^{Aa}	0.64 ± 0.07^{Aa}	271.9 ± 6.9^{Aa}	-0.52 ± 0.06^{Aa}
	A	103 ± 61^{Ab}	32 ± 26^{Aa}	16 ± 13^{Aa}	78.8 ± 20.4^{Ab}	0.69 ± 0.02^{Aa}	270.0 ± 27.1^{Aa}	-0.64 ± 0.02^{Aab}
	C	81 ± 22^{Ab}	25 ± 11^{Aa}	13 ± 6^{Aa}	86.8 ± 3.8^{Ab}	0.46 ± 0.14^{Aa}	290.2 ± 15.8^{Aa}	-0.85 ± 0.13^{Bb}

Capital letters indicate differences across ecosystems within a given horizon (Litter, A, B or C). Ecosystems that do not share the same capital letter can be considered significantly different. Lower case letters indicate differences between horizons within an ecosystem. Horizons that do not share the same lowercase letters can be considered significantly different. Differences of means were determined with mixed effects models using a Bonferroni correction for 3 or 15 tests respectively.

Table S5. Linear regression R^2 results of predictor variables for molecular diversity indices separated by horizon for hydrophilic compounds identified from the LC-MS/MS HILIC column.

Litter					
Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.05	0.01	-0.03	0.03	**0.29****
TN	**0.13 .**	0.05	-0.06	0.06	-0.05
C:N ratio	**0.14 .**	-0.02	-0.06	-0.06	-0.01
MAT (°C)	-0.06	0.06	-0.04	0.10	0.09
MAP (mm)	-0.05	-0.01	0.03	0.10	0.03
Aridity	-0.04	-0.04	0.01	0.07	-0.02
A-horizon					
Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.06	-0.05	-0.06	**0.11 .**	**0.14 .**
TN	-0.05	-0.06	0.03	-0.04	0.02
C:N ratio	-0.06	-0.04	0.00	**0.12 .**	0.08
Fe (mg g^{-1} soil)	0.10	0.01	-0.05	-0.03	-0.05
Al (mg g^{-1} soil)	0.09	0.04	-0.06	-0.06	-0.05
MAT (°C)	-0.05	-0.05	-0.06	0.10	-0.06
MAP (mm)	-0.05	-0.03	-0.06	0.08	-0.06
Aridity	-0.04	-0.02	-0.06	0.01	-0.06
Depth (m)	-0.06	**0.23***	**0.31****	-0.06	-0.06
pH	-0.04	-0.04	0.01	-0.01	-0.05
Silt (%)	**0.13 .**	-0.03	-0.06	-0.01	-0.04
Clay (%)	0.02	-0.07	0.10	0.02	-0.03
B-horizon					
Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{H,q=1}$)**	**Molecular Diversity ($D_{H,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	0.01	0.02	0.14	-0.05	-0.06
TN	0.00	0.02	0.14	0.02	-0.04
C:N ratio	-0.01	-0.03	0.14	0.00	-0.06
Fe (mg g^{-1} soil)	0.06	0.09	0.14	-0.08	-0.06
Al (mg g^{-1} soil)	0.04	0.06	0.14	-0.08	-0.05

MAT (°C)	-0.08	-0.07	0.14	-0.07	-0.05
MAP (mm)	-0.08	-0.08	0.14	-0.05	0.12
Aridity	-0.07	-0.08	0.14	-0.08	0.03
Depth (m)	0.05	0.05	0.14	-0.04	-0.07
pH	**0.14 .**	**0.15 .**	**0.14 .**	-0.06	-0.03
Silt (%)	-0.07	-0.07	0.14	**0.25***	**0.17 .**
Clay (%)	-0.01	0.01	0.14	**0.14 .**	0.02
C-horizon					
Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.06	0.01	0.07	0.04	-0.02
TN	-0.06	0.01	0.08	0.05	0.00
C:N ratio	-0.01	0.04	-0.01	-0.06	-0.01
Fe (mg g^{-1} soil)	0.06	**0.14 .**	**0.13**	-0.05	-0.05
Al (mg g^{-1} soil)	0.01	0.04	0.01	-0.05	-0.06
MAT (°C)	-0.05	0.02	0.05	-0.03	**0.12 .**
MAP (mm)	-0.02	0.01	-0.02	-0.06	0.01
Aridity	0.00	0.04	0.02	-0.06	-0.01
Depth (m)	-0.01	-0.06	-0.05	0.00	**0.21***
pH	-0.06	-0.04	0.00	0.02	0.04
Silt (%)	-0.05	-0.07	-0.07	0.04	-0.07
Clay (%)	0.10	0.03	-0.02	0.05	0.02

Iron and aluminum concentrations are from hydroxylamine HCl extracts. Significance of fit signified by p-values are reported as asterisks: p-value < 0.001 ***, p-value < 0.01 **, p-value < 0.05 *, and p-value < 0.10 are denoted by (.).

Table S6. Linear regression R^2 results of predictor variables for molecular diversity indices separated by horizon for hydrophobic compounds from the LC-MS/MS C18 column.

Litter

Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	**0.12 .**	-0.03	-0.06	-0.03	**0.21***
TN	0.02	0.04	-0.04	-0.06	-0.06
C:N ratio	**0.26***	0.05	-0.04	-0.01	0.02
MAT (°C)	**0.17***	**0.15 .**	0.09	-0.06	0.02
MAP (mm)	**0.18***	-0.05	-0.02	-0.02	**0.18***
Aridity	**0.10 .**	-0.06	0.01	0.00	**0.11 .**

A-horizon

Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.05	-0.04	-0.02	-0.06	0.04
TN	-0.04	-0.02	0.00	-0.06	-0.03
C:N ratio	-0.04	-0.04	-0.04	-0.05	0.00
Fe (mg g^{-1} soil)	**0.21***	**0.26***	**0.22***	-0.06	0.03
Al (mg g^{-1} soil)	0.07	0.05	0.01	-0.06	**0.27***
MAT (°C)	-0.01	0.02	0.04	-0.06	**0.11 .**
MAP (mm)	0.08	0.02	-0.01	-0.05	**0.36****
Aridity	**0.16 .**	**0.11 .**	0.07	-0.05	**0.25***
Depth (m)	-0.04	-0.04	-0.05	**0.10 .**	-0.05
pH	**0.12 .**	0.00	-0.04	0.02	0.05
Silt (%)	0.00	0.10	**0.12 .**	0.07	-0.04
Clay (%)	**0.22***	0.06	-0.02	0.09	**0.21***

B-horizon

Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{H,q=1}$)**	**Molecular Diversity ($D_{H,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.04	-0.07	-0.06	-0.07	-0.07
TN	-0.05	-0.07	-0.06	-0.05	-0.08
C:N ratio	-0.04	-0.07	-0.06	-0.06	0.06
Fe (mg g^{-1} soil)	-0.08	-0.07	-0.08	-0.07	-0.06
Al (mg g^{-1} soil)	-0.06	-0.08	-0.08	-0.08	-0.04
MAT (°C)	0.09	0.06	0.00	-0.07	-0.08

MAP (mm)	-0.01	-0.07	-0.06	-0.08	-0.08
Aridity	0.00	-0.06	-0.05	-0.08	-0.06
Depth (m)	0.06	0.02	0.02	0.13	-0.05
pH	**0.38****	**0.21***	**0.15 .**	-0.06	-0.07
Silt (%)	0.04	**0.17 .**	**0.19 .**	-0.06	0.07
Clay (%)	0.07	0.11	0.10	-0.03	0.03
C-horizon					
Predictor	**Molecular Richness (D_R)**	**Molecular Diversity ($D_{HN,q=1}$)**	**Molecular Diversity ($D_{HN,q=2}$)**	**Functional Molecular Diversity ($D_{Rao(NOSC)}$)**	**Functional Molecular Diversity ($D_{Rao(MW)}$)**
SOC	-0.06	-0.04	-0.02	0.13	-0.06
TN	-0.06	-0.03	0.00	0.12	-0.05
C:N ratio	-0.02	-0.03	-0.01	-0.05	-0.03
Fe (mg g^{-1} soil)	0.08	**0.14 .**	**0.17***	-0.06	-0.06
Al (mg g^{-1} soil)	**0.15 .**	**0.14 .**	**0.16 .**	0.03	-0.04
MAT (°C)	0.07	**0.19***	**0.23***	0.04	0.02
MAP (mm)	0.05	-0.05	-0.05	0.08	-0.02
Aridity	0.10	-0.02	-0.03	0.03	-0.03
Depth (m)	-0.06	0.01	0.00	-0.05	-0.04
pH	0.07	0.01	0.00	-0.05	-0.04
Silt (%)	-0.07	-0.06	-0.07	-0.04	-0.02
Clay (%)	-0.05	-0.06	-0.04	-0.01	**0.22***

Iron and aluminum concentrations are from hydroxylamine HCl extracts. Significance of fit signified by p-values are reported as asterisks: p-value < 0.001 ***, p-value < 0.01 **, p-value < 0.05 *, and p-value < 0.10 are denoted by (.).

Table S7. Mixed effect model results to test the significance of depth on molecular diversity of the hydrophilic compounds from the LC-MS/MS HILIC column.

Response Variable	**F-statistic**	**Degrees of freedom**	**p-value**
Molecular Richness (D_R)			
Depth	11.79	1	0.001
Ecosystem Type	0.64	5	0.67
Molecular Diversity ($D_{H,q=1}$)			
Depth	7.94	1	0.007
Ecosystem Type	0.77	5	0.57
Molecular Diversity ($D_{H,q=2}$)			
Depth	9.17	1	0.003
Ecosystem Type	0.85	5	0.52
Functional Molecular Diversity ($D_{Rao(NOSC)}$)			
Depth	0.02	1	0.90
Ecosystem Type	1.78	5	0.13
Functional Molecular Diversity ($D_{Rao(MW)}$)			
Depth	1.21	1	0.28
Ecosystem Type	1.48	5	0.21

Mixed effects models included ecosystem type and depth (m) as fixed effects and sample location as a random effect. F-statistics are shown alongside degrees of freedom and p-values.

Table S8. Mixed effect model results to test the significance of depth on molecular diversity of the hydrophobic compounds from the LC-MS/MS C18 column.

Response Variable	**F-statistic**	**Degrees of freedom**	**p-value**
Molecular Richness (D_R)			
Depth	18.99	1	<0.0001
Ecosystem Type	0.92	5	0.47
Molecular Diversity ($D_{H,q=1}$)			
Depth	17.37	1	0.0002
Ecosystem Type	0.64	5	0.67
Molecular Diversity ($D_{H,q=2}$)			
Depth	13.11	1	0.0006
Ecosystem Type	1.08	5	0.38
Functional Molecular Diversity ($D_{Rao(NOSC)}$)			
Depth	3.59	1	0.06
Ecosystem Type	0.42	5	0.83
Functional Molecular Diversity ($D_{Rao(MW)}$)			
Depth	1.56	1	0.22
Ecosystem Type	0.53	5	0.75

Mixed effects models included ecosystem type and depth (m) as fixed effects and sample location as a random effect. F-statistics are shown alongside degrees of freedom and p-values.

Table S9. Mixed effect model results to test the significance of ecosystem type on molecular diversity of the hydrophilic compounds from the LC-MS/MS HILIC column.

Response Variable	**Null Model AIC**	**Model AIC**	**Chi-squared**	**p-value**
Molecular Richness (D_R)	944.37	952.65	11.72	0.30
Molecular Diversity ($D_{H,q=1}$)	664.55	676.69	7.86	0.64
Molecular Diversity ($D_{H,q=2}$)	553.33	566.26	7.07	0.72
Functional Molecular Diversity ($D_{Rao(NOSC)}$)	-34.33	--31.37	17.03	0.07
Functional Molecular Diversity ($D_{Rao(MW)}$)	715.26	719.29	16.05	0.10

Null models include depth (m) as a fixed effect and sample location as a random effect. Models with ecosystem type add ecosystem type as a fixed effect term. AIC is the Akaike information criterion. Chi-squared and associated p-values from likelihood ratio tests comparing null model to the model containing ecosystem type.

Table S10. Mixed effect model results to test the significance of ecosystem type on molecular diversity hydrophobic compounds from the LC-MS/MS C18 column.

Response Variable	**Null Model AIC**	**Model AIC**	**Chi-squared**	**p-value**
Molecular Richness (D_R)	852.25	857.69	14.63	0.15
Molecular Diversity ($D_{H,q=1}$)	652.82	660.87	11.95	0.29
Molecular Diversity ($D_{H,q=2}$)	531.81	539.12	12.69	0.24
Functional Molecular Diversity ($D_{Rao(NOSC)}$)	-76.97	-65.42	8.45	0.59
Functional Molecular Diversity ($D_{Rao(MW)}$)	669.73	681.86	7.87	0.64

Null models include depth (m) as a fixed effect and sample location as a random effect. Models with ecosystem type add ecosystem type as a fixed effect term. AIC is the Akaike information criterion. Chi-squared and associated p-values from likelihood ratio tests comparing null model to the model containing ecosystem type.

SI References

1. D. H. Wall, U. N. Nielsen, J. Six, Soil biodiversity and human health. *Nature* **528**, 69–76 (2015).
2. M. Lange, *et al.*, Plant diversity increases soil microbial activity and soil carbon storage. *Nat. Commun.* **6**, 1–8 (2015).
3. F. Bastida, *et al.*, Soil microbial diversity–biomass relationships are driven by soil carbon content across global biomes. *ISME J.* **15**, 2081–2091 (2021).
4. S. Naeem, Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems. *Phil. Trans. Royal Soc. Lon.* B **347**, 249–262 (1995).
5. 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).
6. M. Zark, T. Dittmar, Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 1–8 (2018).
7. 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).
8. 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).
9. A. J. Daly, J. M. Baetens, B. De Baets, Ecological diversity: Measuring the unmeasurable. *Mathematics* **6** (2018).
10. L. Jost, Entropy and diversity. *Opinion* **2**, 363–375 (2006).
11. 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).
12. A. Chao, *et al.*, Rarefaction and extrapolation with Hill Numbers: A framework for sampling and estimation in species diversity Sstudies. *Ecol. Monogr.* **84**, 45–67 (2014).
13. D. Schleuter, M. Daufresne, F. Massol, C. Argillier, A user's guide to functional diversity indices. *Ecol. Monogr.* **80**, 469–484 (2010).
14. 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).
15. Anne Chao, Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**, 265–270 (1984).
16. A. Chao, Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987).
17. C. H. Chiu, A. Chao, Distance-based functional diversity measures and their decomposition: A framework based on hill numbers. *PLoS One* **9** (2014).
18. Z. (Sam) Ma, L. Li, Measuring metagenome diversity and similarity with Hill numbers. *Mol. Ecol. Resour.* **18**, 1339–1355 (2018).
19. Z. Botta-Dukat, Rao's quadratic entropy as a measure of functional diversity based

on multiple traits. *J. Veg. Sci.* **16**, 533–540 (2005).

20. X. M. Li, *et al.*, Organic carbon amendments affect the chemodiversity of soil dissolved organic matter and its associations with soil microbial communities. *Environ. Sci. Technol.* **53**, 50–59 (2019).
21. X. M. Li, *et al.*, Molecular chemodiversity of dissolved organic matter in paddy soils. *Environ. Sci. Technol.* **52**, 963–971 (2018).
22. C. Ricotta, L. Szeidl, Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. *Theor. Popul. Biol.* **70**, 237–243 (2006).
23. H. Peter, *et al.*, Function-specific response to depletion of microbial diversity. *ISME J.* **5**, 351–361 (2011).