Supplementary discussion S1. We observed a cluster of unsaturated compounds with low O/C ratio which were predominantly positively linked to biomass-normalized respiration but negatively or not linked to biomass-normalized growth. A common example for unsaturated hydrocarbons are alkenes, however those are thought to be comparatively easy to metabolize. Possibly, the unsaturated hydrocarbon compounds in question could stem from cutan and suberan, which can derive from the decomposition of abundant lipid plant biopolymers such as cutin (primarily leaf- derived) and suberin (primarily root-derived), and are hard to decompose for microbes [1]. While cutan and suberan are in principal saturated, it has been found that they can have minor contributions of unsaturated structures [2,3] which could place them in the range of unsaturated compounds with very low O/C ratio. The association of cutan and suberan with the cluster of unsaturated compounds with very low O/C ratio is speculative, but future MS/MS experiments might help to elucidate the links between metabolism and potential suberin/cutin substructures.

References supplementary discussion

1. Kögel-Knabner, I. & Amelung, W. Dynamics, Chemistry, and Preservation of Organic Matter in Soils. in *Treatise on Geochemistry* 157–215 (Elsevier, 2014). doi:10.1016/B978-0-08-095975-7.01012-3.
2. Turner, J. W., Hartman, B. E. & Hatcher, P. G. Structural characterization of suberan isolated from river birch (Betula nigra) bark. *Org. Geochem.* **57**, 41–53 (2013).
3. Leide, J. *et al.* Leaf cuticle analyses: implications for the existence of cutan/non-ester cutin and its biosynthetic origin. *Ann. Bot.* **126**, 141–162 (2020).

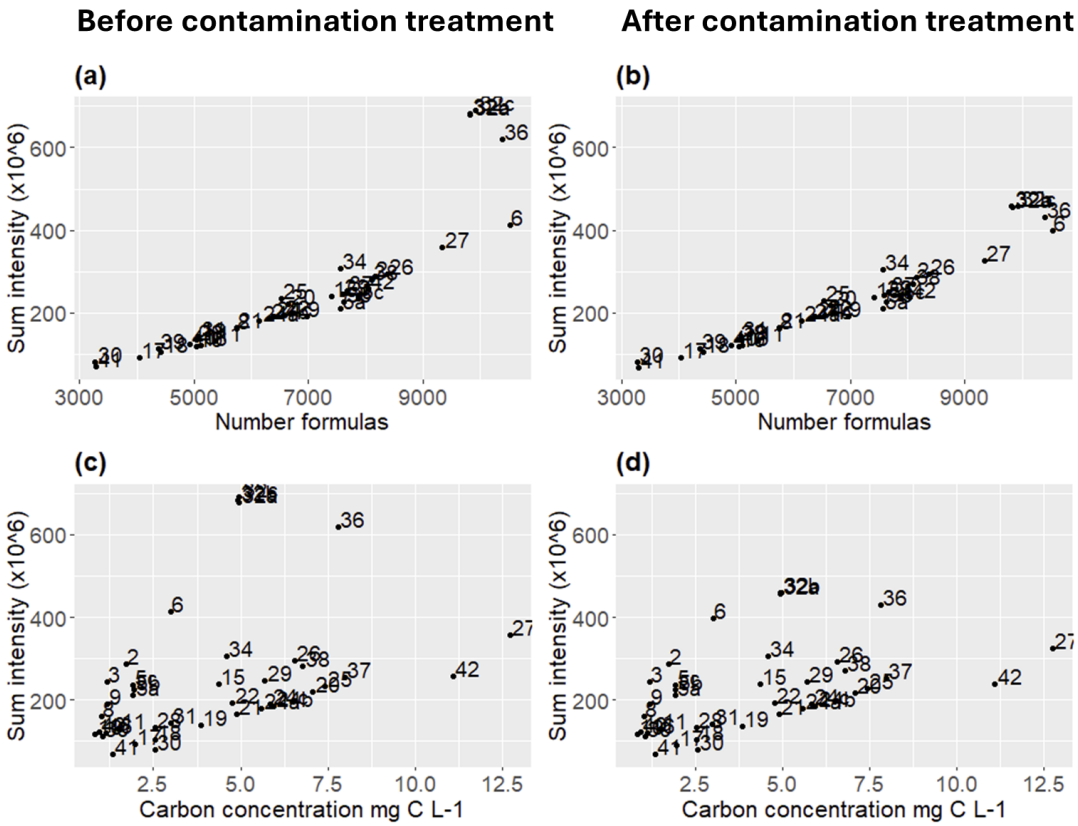

Figure S1. Total signal intensity before and after outlier treatment. Panels (a) and (b) show total signal intensity against the number of formulas per sample, and panels (c) and (d) against organic carbon concentration. Numbers indicate the sample ID.

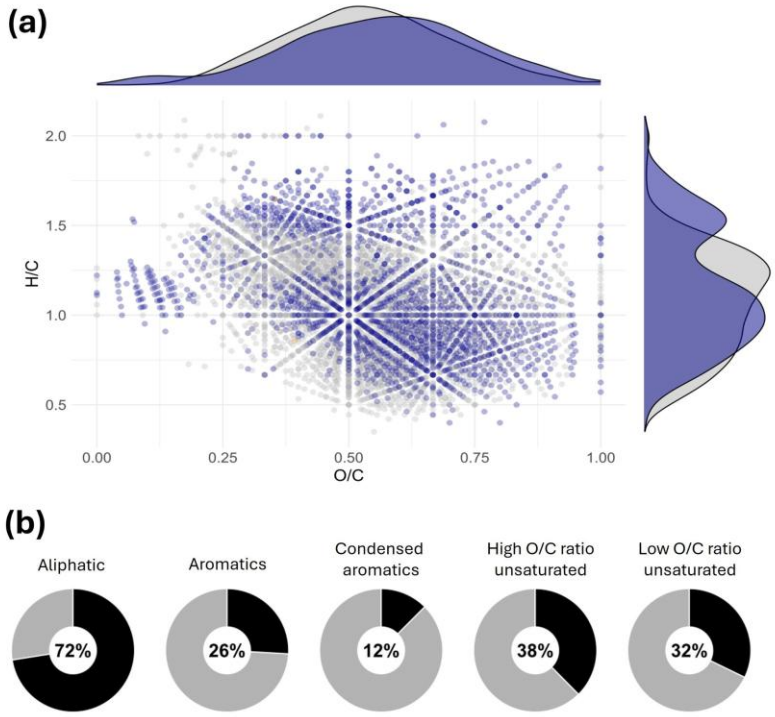

Figure S2. Formulas significantly correlated (adjusted p-value < 0.05) with organic carbon concentration in extracts before normalization. Panel (a) shows a Van Krevelen diagram (dark blue = positive correlation, light orange = negative correlation), panel (b) shows the percentage of correlated formulas in each chemical class. We performed Pearson correlation and adjusted p-values with the Benjamini-Hochberg method to correct for the false discovery rate.

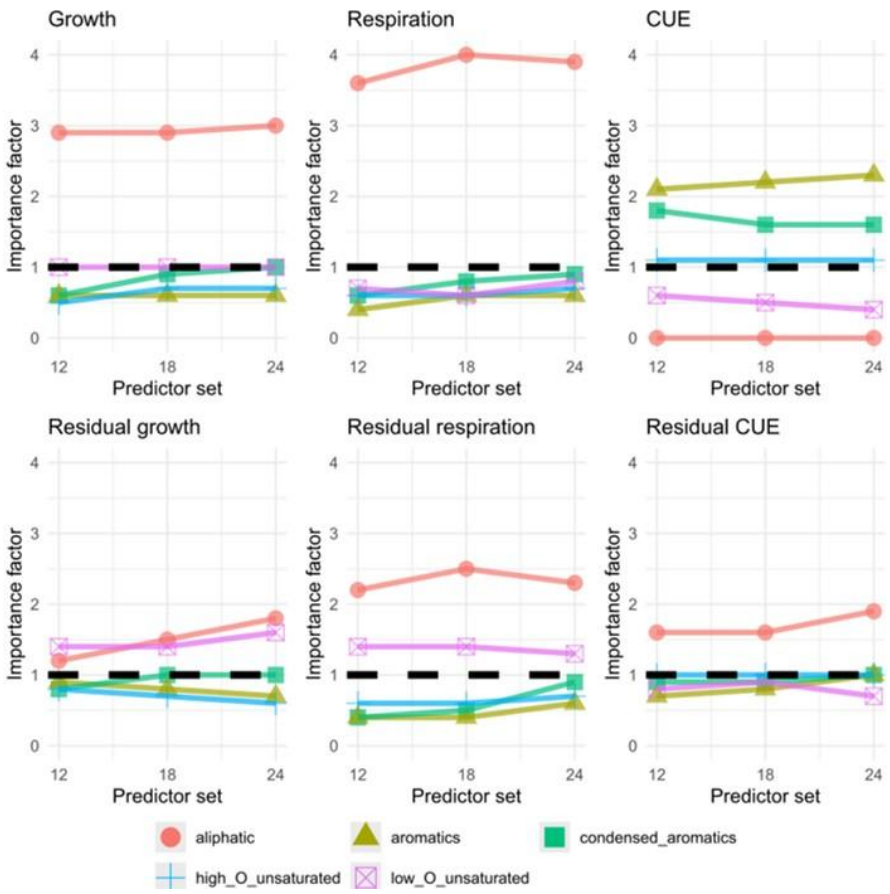

Figure S3. Sensitivity analysis to assess whether the study results are affected by the selection of formulas included in the PLS regression. Each panel shows the results for one response variable. The x-axis shows results for models that consider all molecular formulas that occurred in >11, >17 and >23 soils, with the 10% most important formulas retained. "growth" and "respiration" always refer to biomass-normalized rates.

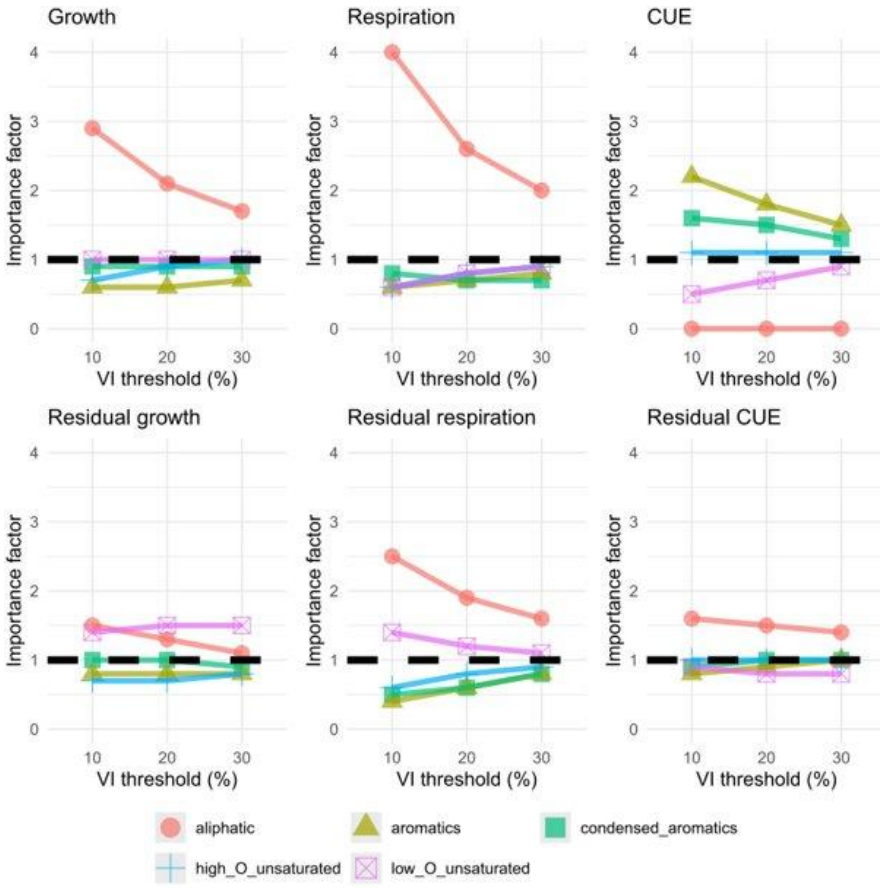

Figure S4. Sensitivity analysis to assess whether the study results depend on the threshold used to select predictor variables based on their variable importance (VI). Each panel shows the results for one response variable. The x-axis shows results when the 10%, 20% and 30% most important variables are interpreted, based on models that considered molecular formulas that occurred in >17 soils. "growth" and "respiration" always refer to biomass-normalized rates.

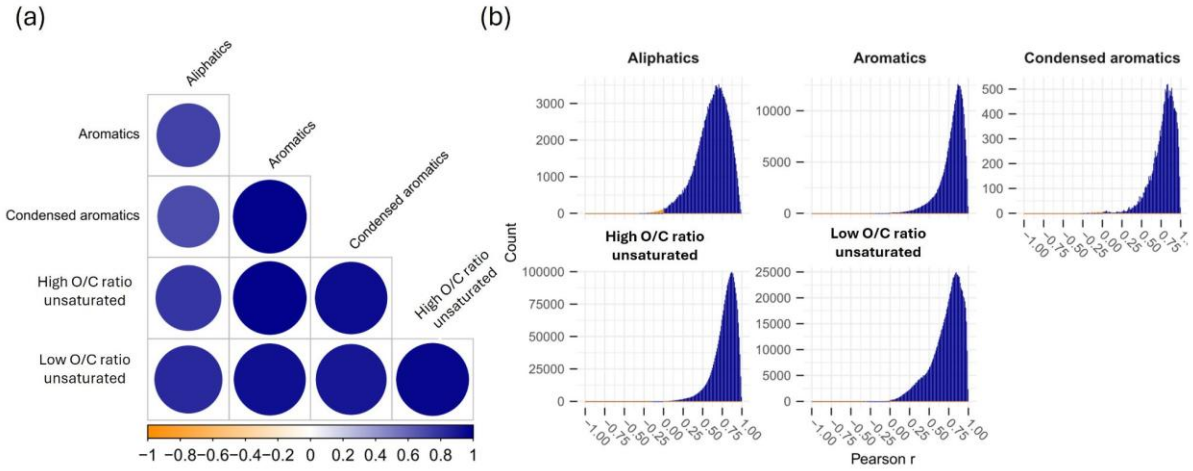

Figure S5. Correlation among normalized peak intensities across chemical classes for the subset of the mass spectra that was used for PLS regression. Panel (a) shows significant correlations (Pearson correlation, $p < 0.05$) between mean normalized peak intensities of the different chemical classes. Circle size and color indicate the strength of the relationships (Pearson correlation coefficient). Panel (b) shows the distribution of pairwise Pearson correlation coefficients across all compounds within each chemical class.

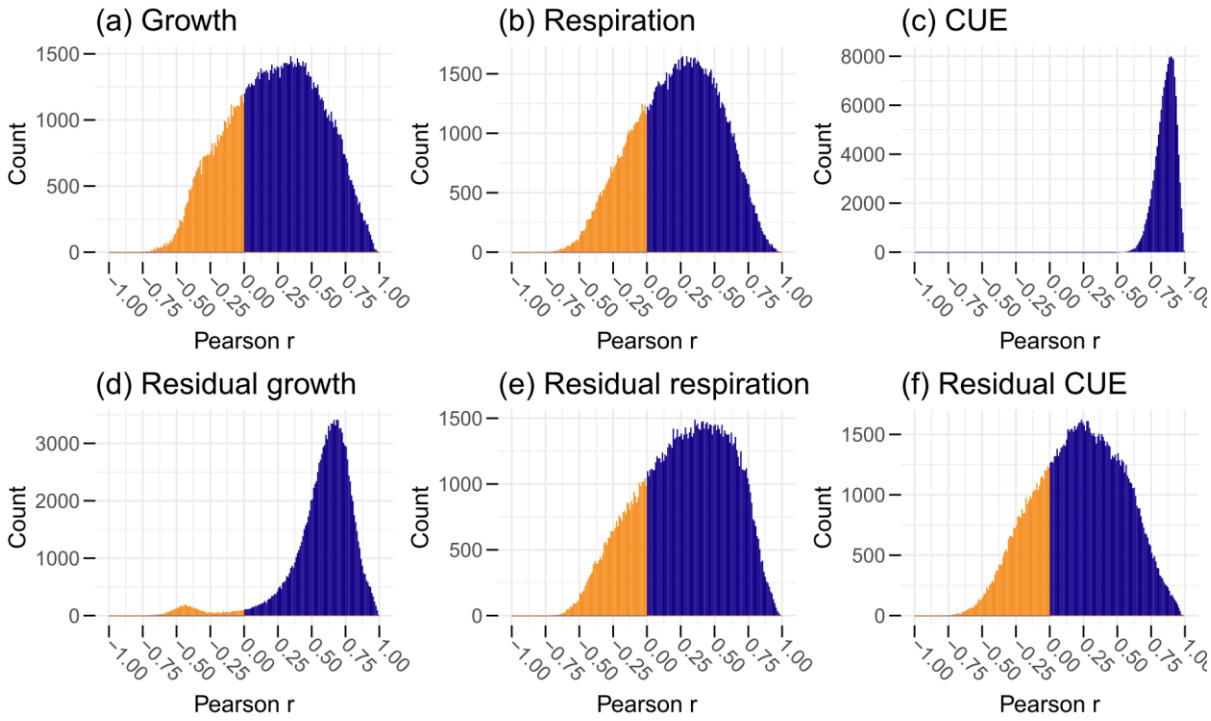

Figure S6. Correlation among normalized peak intensities of the top 10% most important predictive molecular formulas in each model. Shown are the distributions of pairwise Pearson correlation coefficients. The upper panels (a) to (c) show models which were fitted to the full variation of growth, respiration and carbon use efficiency (CUE), while the lower panels (d) to (f) show the models which were fitted only to the residual variations from a previous study [11]. "growth" and "respiration" always refer to biomass-normalized rates.

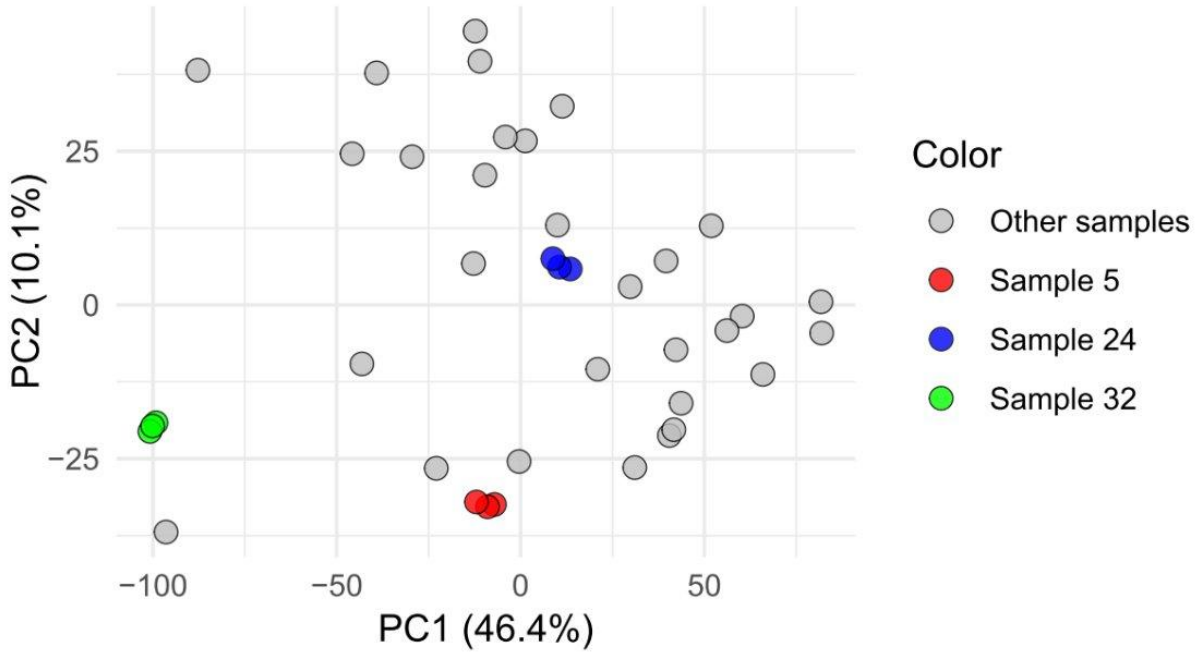

Figure S7. The first two principal components (together 56.5 % variation) based on the scaled signal intensities of the 5305 compounds that were retained for PLS regression. Each dot represents one sample, highlighted in colors are the three triplicates. This plot shows that the chosen methodology to characterize the chemical composition of extracted organic matter is highly reproducible and sensitive enough to reliably distinguish different samples.

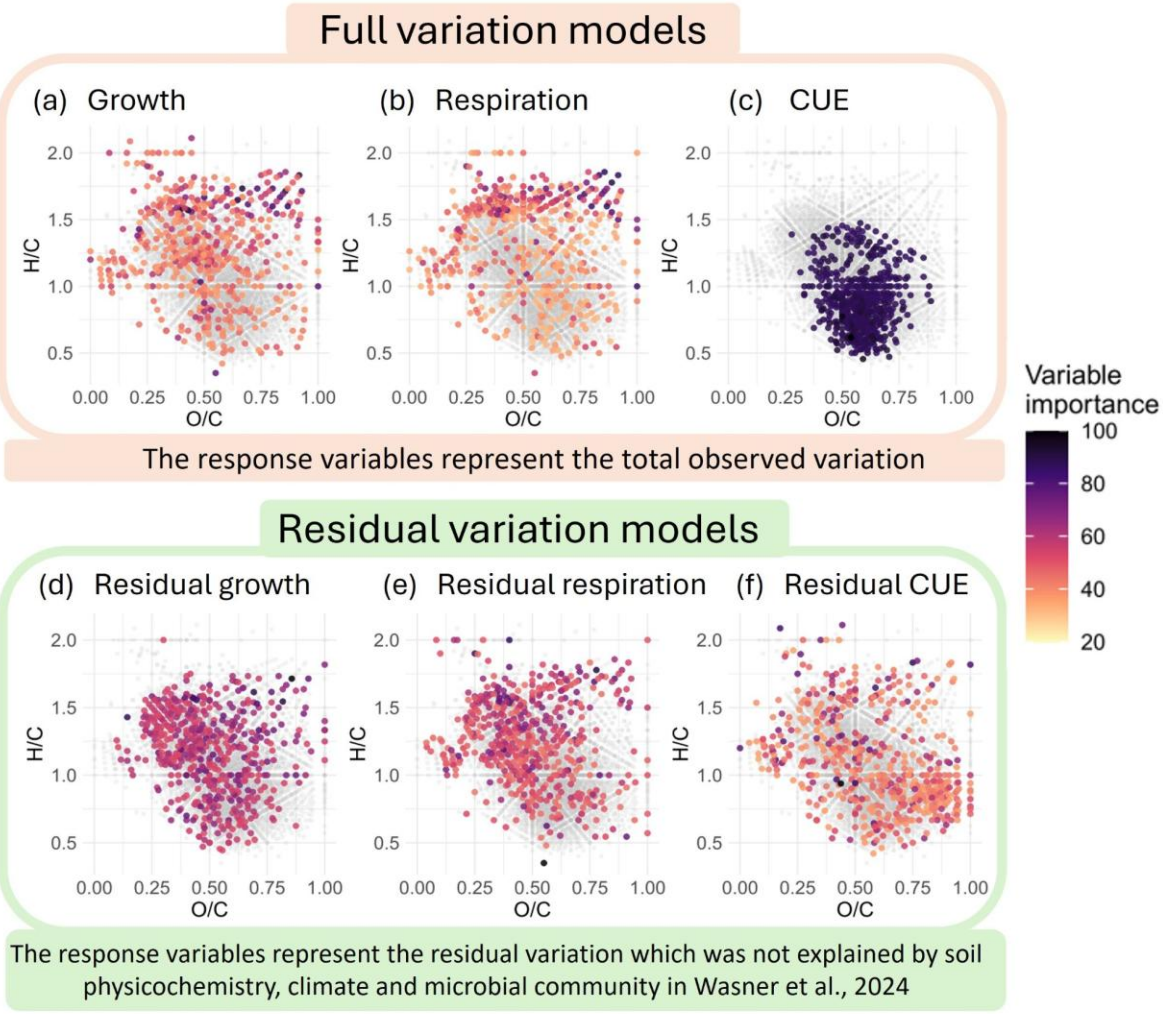

Figure S8. Van Krevelen diagrams showing the most important predictive molecular formulas for each of the six models. Shown in grey are all molecular formulas, highlighted in color are the top 10% most important formulas. Coloration indicates the variable importance with which molecular formulas feature in the models. The upper panels (a) to (c) show models which were fitted to the full variation of growth, respiration and carbon use efficiency (CUE), while the lower panels (d) to (f) show the models which were fitted only to the residual variations from a previous study [11]. "growth" and "respiration" always refer to biomass-normalized.

Table S1. Summary of the chemical composition of extractable organic matter. Shown are minimum (min.), mean, median and maximum (max.) values across the 33 investigated soils, for all chemical classes together as well as for the individual chemical classes. Shown is data from the entire mass spectra (i.e., including rare molecular formulas not considered for PLS regression). °Percent of summed signal intensities per sample; *Intensity-weighted mean. NOSC = Nominal oxidation state of carbon.

Chemical class	Parameter	Min.	Mean	Median	Max.
All	Number formulas	3275	6855	6650	12473
	Signal intensity° (%)	100.0	100.0	100.0	100.0
	Mean NOSC*	-0.22	0.04	0.05	0.27
High O/C ratio Unsaturated	Number formulas	934	2759	2783	5636
	Signal intensity° (%)	29.9	45.6	46.1	59.7
	Mean NOSC*	0.31	0.38	0.38	0.47
Low O/C ratio Unsaturated	Number formulas	871	1704	1742	3164
	Signal intensity° (%)	18.2	27.2	26.6	34.1
	Mean NOSC*	-0.54	-0.44	-0.44	-0.36
Aliphatics	Number formulas	635	1080	941	2103
	Signal intensity° (%)	4.6	13.5	12.2	29.0
	Mean NOSC*	-1.02	-0.66	-0.65	-0.42
Aromatics	Number formulas	354	986	943	1940
	Signal intensity° (%)	8.5	11.3	10.9	18.4
	Mean NOSC*	0.38	0.52	0.53	0.60
Condensed Aromatics	Number formulas	114	326	261	984
	Signal intensity° (%)	1.4	2.4	2.0	6.3
	Mean NOSC*	0.67	0.72	0.71	0.79

Table S2. Summary of linear regression with intensity-weighted NOSC for all six response variables. All models have 31 degrees of freedom. transf. = response variable transformed with natural logarithm; coef. = coefficient. "growth" and "respiration" always refer to biomass-normalized rates.

Variable	transf.	p-value	adj. R2	coef.
Growth	yes	0.71	-0.03	0.28
Respiration	yes	0.07	0.07	-1.82
CUE	**yes**	**0.03**	**0.12**	**1.49**
Residual growth	no	0.36	0	0.45
Residual respiration	no	0.45	-0.01	-0.34
Residual CUE	no	0.78	-0.03	-0.03

Table S3. Results from two redundancy analyses to find links between organic matter composition and (a) environmental factors (soil physicochemistry & climate) and (b) microbial community composition. Predictor variables were taken from Wasner et al. (2024, Global Change Biology). Shown are significant predictors (p-value < 0.05), and percentages of total variation explained by the predictors. The "PCs" in panel (b) refer to different dimensions of microbial community composition which are explained in Wasner et al. (2024, Global Change Biology). Abbreviations: MAT = Mean annual temperature; OM = organic matter; var. = variation.

(a) RDA with environmental condition

Predictor	Percent of total var. explained	Cumulative percent of total var. explained	P-value
MAT	11.2	11.2	0.001
OM Quantity	7.9	19.1	0.014
Sand	3.8	22.9	0.031
pH	2.4	25.3	0.038

(b) RDA with microbial community composition

Predictor	Percent of total var. explained	Cumulative percent of total var. explained	P-value
PC1	17.3	17.3	0.001
PC4	9.5	26.8	0.002
PC2	6.5	33.3	0.001
PC5	3.0	36.2	0.014
PC8	2.7	38.9	0.028
PC3	1.4	40.4	0.035