SUPPLEMENTARY INFORMATION

Solid phase molecular diversity enhances soil organic matter persistence

Zhen Yang[1], Lin Zhang[2], Siqi Gao[1,3], Songlin Wu[1,3], Ke-Qing Xiao[1,3], Hanyong Peng[3,4], Tianran Sun[1,3*], Hailiang Dong[5], Xing-Guo Han[6,7], Andreas Kappler[8,9], Johannes Lehmann[10,11,12,13], Yong-Guan Zhu[1,3]

1. State Key Laboratory of Regional and Urban Ecology, Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China.
2. China Innovation Center, Shimadzu (China) Co. LTD. Beijing Branch, Beijing, 100020, China
3. University of Chinese Academy of Sciences, Beijing 100049, China.
4. State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China.
5. Center for Geomicrobiology and Biogeochemistry Research, State Key Laboratory of Geomicrobiology and Environmental Changes, China University of Geosciences, Beijing, China.
6. School of Life Sciences, Hebei University, Baoding, 071002, China.
7. State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China.
8. Geomicrobiology, Department of Applied Geosciences, Tuebingen University, 72076, Germany.
9. Cluster of Excellence: EXC 2124: Controlling Microbes to Fight Infection, Tuebingen, 72076, Germany.
10. Soil and Crop Sciences, School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14850, USA.
11. Department of Global Development, Cornell University, Ithaca, NY 14850, USA.
12. Cornell Institute for Digital Agriculture, Cornell University, Ithaca, NY 14850, USA.
13. Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY 14850, USA.
* Corresponding author; Email: trsun@rcees.ac.cn

Number of figures in supporting information: 22

Number of tables in supporting information: 2

Total numbers of pages in supporting information: 30 (including cover page)

Section 1: Supplementary methods

1. Analysis of molecular richness and diversity. Molecular diversity of POM and MAOM was followed by the same method published by Davenport (2023). We calculated 1) Molecular richness (D_R) using the sum of molecules from identified peaks indicating by mass-to-charge ratio (m/z) in each sample; 2) Molecular diversity using Hill numbers (D_{HN}, q=2). Molecular richness and relative peak intensity are considered. Hill numbers represent a unifying diversity index that incorporates molecular richness and evenness (Supplementary Table S1).

2. Proportion calculation. The proportions of molecular richness and intensity-based molecular diversity were ascertained by calculating the ratio of molecular richness or molecular diversity within a 1 kDa range to the aggregate molecular richness or molecular diversity (the sum across four soil layers within the 0-10 kDa molecular weight range) for each sample.

$$D_{Rpi} = \frac{D_{Ri}}{D_{Rtot}} \quad \text{..... .... (Eq. 1)}$$

$$D_{Hpi} = \frac{D_{Hi}}{D_{Htot}} \quad \text{..... ..... (Eq. 2)}$$

Where D_{Ri} and D_{Hi} are molecular richness or diversity within each 1 kDa (e.g. i = 0-1, 1-2, 2-3,..... or 9 - 10 kDa). D_{Rtot} and D_{Rtot} are molecular richness or diversity within 0-10 kDa.

3. β calculation. The relation between $d^{13}C$ and the logarithm in base 10 of C content (log SOC) was investigated by linear regression for each soil. The slope of the regression is defined as the β value. The power functions associated with the decrease of C content or to the increase of $d^{13}C$ value with depth were fitted in STATISTICA 10.0 using the following equation:

$$C_d = C_{Ah} d^k \quad \text{..... ....(Eq. 3)}$$

$$\delta^{13}C_d = \delta^{13}C_{Ah} d^l \quad \text{..... ..... (Eq. 4)}$$

where C_d, $d^{13}C_d$, C_{Ah}, and $d^{13}C_{Ah}$ are the C contents and the $d^{13}C$ value estimated for the depth d and measured in the Ah horizon, respectively, d is the depth in cm, and k and l are the fitted parameters of the functions.

The vertical changes in $\delta^{13}C$ along soil profiles have been proven to be related to soil C turnover (Powers & Schlesinger 2002). The slope of a linear regression relating soil $\delta^{13}C$ to

the log-transformed C concentration was used to describe the SOC turnover rate at different spatial scales (Golubsov et al. 2022; Wang et al. 2017).

$$\delta^{13}\mathrm{C} = \beta \log(\mathrm{C}) + \mathrm{b} \quad \text{..... .....(Eq. 5)}$$

The slope of the regression (β) is a proxy of the SOC turnover rate; a lower β (steeper slope) is associated with a faster turnover rate of soil C, and b is a constant.

4. Redundancy analysis (RDA). RDA is a method of ordination that has evolved from correspondence analysis (CA), combining correspondence analysis with multiple regression analysis. Each step of the calculation is regressed with environmental factors, also known as multivariate direct gradient analysis. RDA is based on linear model redundancy analysis, which is an ordination method that combines regression analysis with principal component analysis, and is an extension of multiple response variable (multiresponse) regression analysis. Conceptually, RDA is the principal CA analysis of the fitted value matrix of the multiple multiple linear regression between the response variable matrix and the explanatory variable matrix. It examines the distribution of each variable in the explanatory and response matrices, as well as the relationship diagrams of each variable with itself and any other variables in any other matrices.

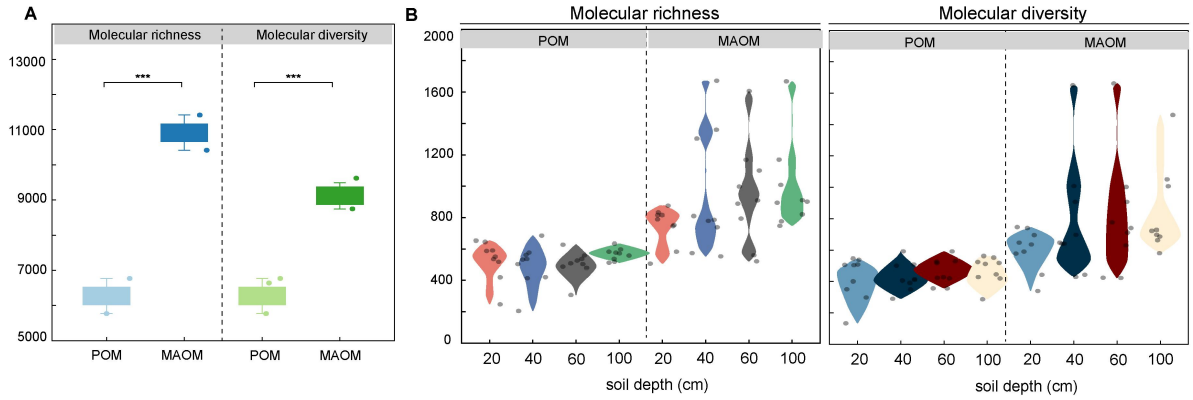
5. Background treatments. To clarify the diversity characterization result from organic matter in soil samples, we burned POM and MAOM samples (at 900°C), which have removed organic carbon from the samples. The control treatment result showed that the range of these peaks of the burned samples was entirely distinct from those of the POM and MAOM samples (Supplementary Fig. 18). Additionally, burned samples showed significantly ($p < 0.001$) lower molecular richness and diversity than POM and MAOM samples (Supplementary Fig. 19). This control result indicated that organic matter predominantly contributed to a high molecular richness and diversity of solid-phase organic fractions measured in this study.

6. Control treatments using mixture of bovin serum albumin (BSA) powder and soil samples. Validation control experiments using BSA powder (MW=6.6 kD) as standard, the homogeneous mixture of POM or MAOM samples and BSA (1:50, m/m) was deposited onto a carbon tape-coated polymeric MALDI target plates to detect m/z peak position in the range from 5-10 kDa (Supplementary Fig. 20). The main peak position located at 6.6 kDa for mixture samples is consistent with the peak position of pure BSA powder, which ensure detection validation.

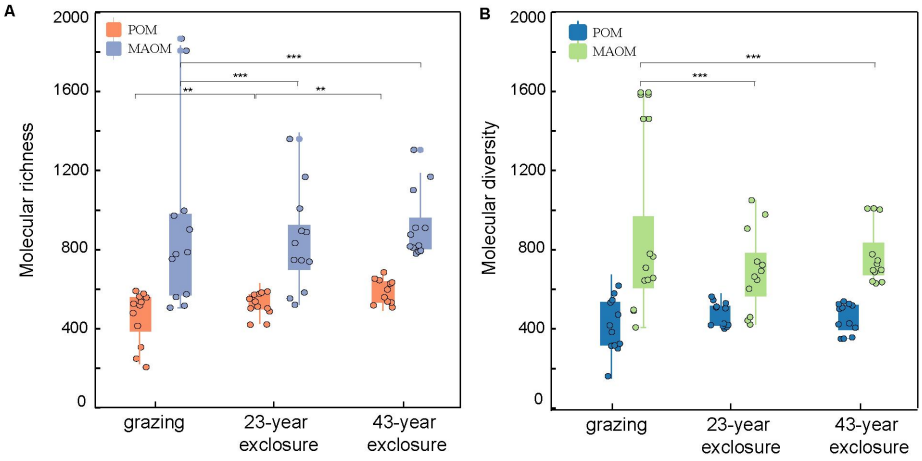
7. Control treatments using standard humic substances. The dissolved fraction and initial powder of IHSS Pahokee Peat Humic Acid (PPHA, standard 1S103H) in the m/z range of 0-2 kDa was detected following the previous method using MALDI-TOF-MS spectrometry (Nicolau et al., 2015) and following the method in this study using carbon tape (NEM, FN731-5N)-coated with polymeric MALDI target plates, respectively. The number of peaks is higher in initial PPHA powder than that in dissolved PPHA extracts (Supplementary Fig. 21A), the comparable abundance of soil composition is obtained by comparing the relative abundance of soil components between dissolved extracts following previous method and the original PPHA powder (Supplementary Fig. 21B). These control experiment results confirmed that effective detection of solid-phase organic fractions by MALDI-TOF-MS spectrometry.

Section 2: Supplementary figures and tables

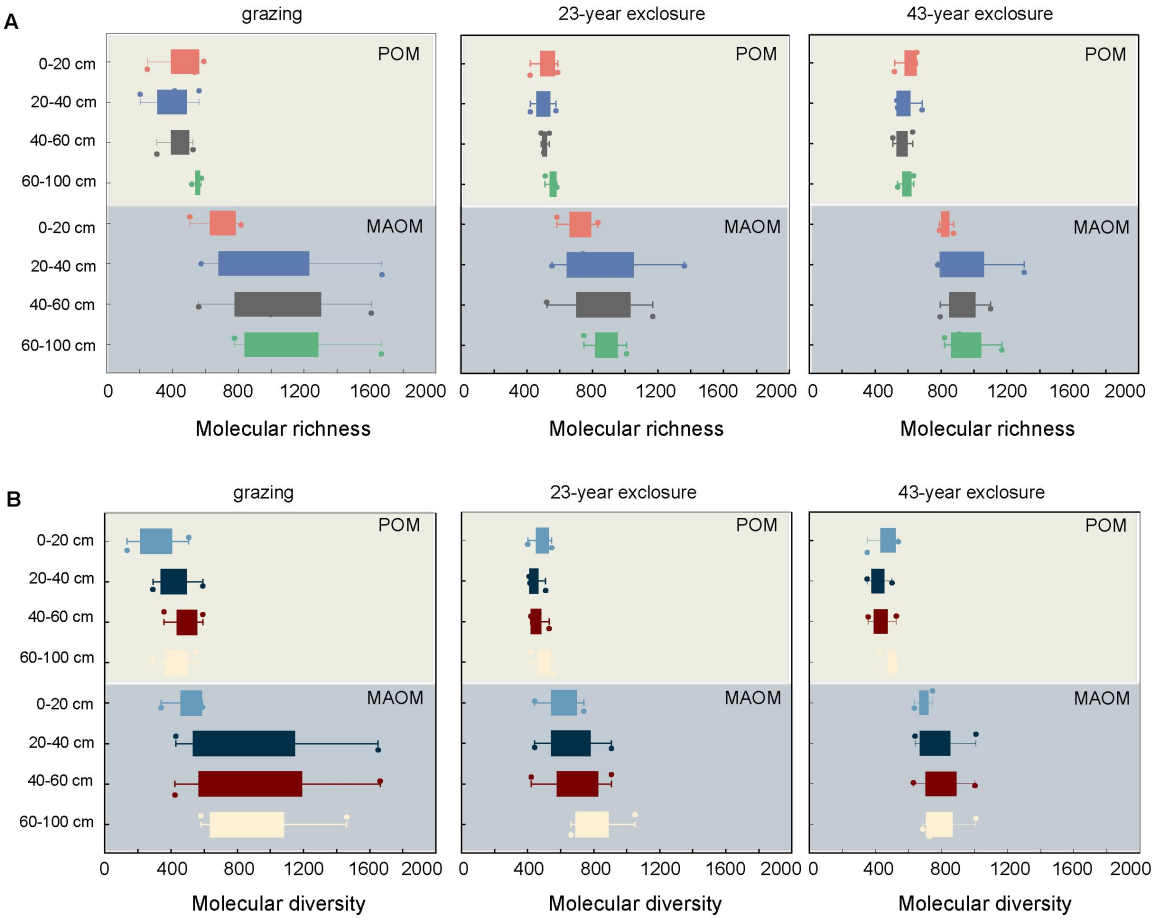
Figures:



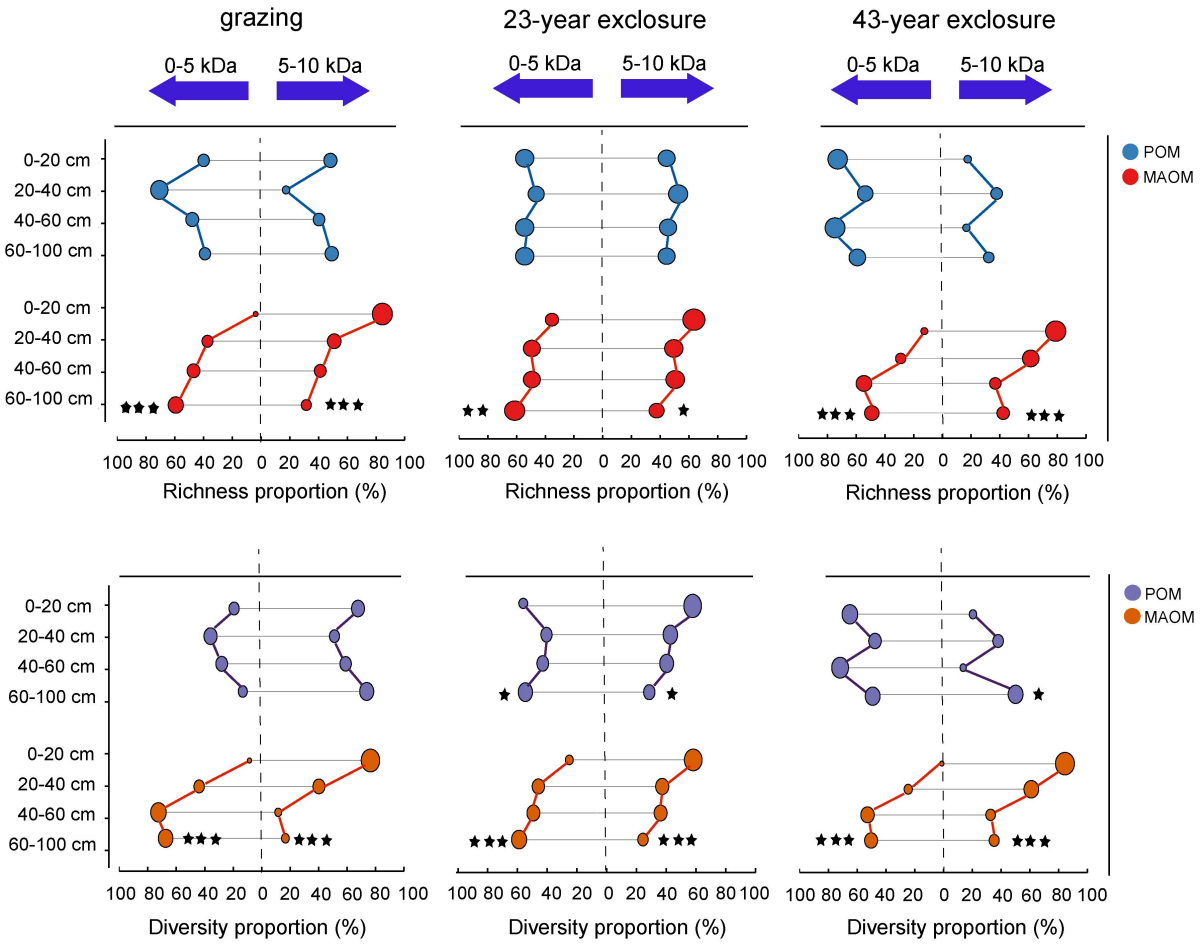
Supplementary Fig. 1. (A). Molecular richness and richness proportion of POM and MAOM, respectively. * indicates $p < 0.05$, **indicates $p < 0.01$, *** indicates $p < 0.001$. (B) Molecular richness and richness of POM and MAOM along the soil profiles.



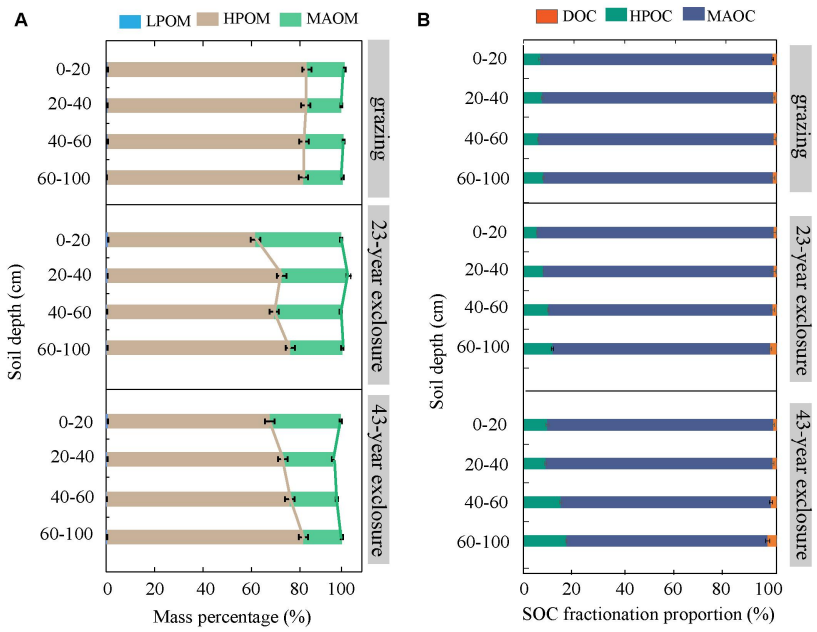
Supplementary Fig. 2. Molecular richness and diversity of POM and MAOM in the grazing, 23-year, and 43-year exclosure treatment, respectively. * indicates $p < 0.05$, **indicates $p < 0.01$, *** indicates $p < 0.001$. Not mentioned implies no significance.



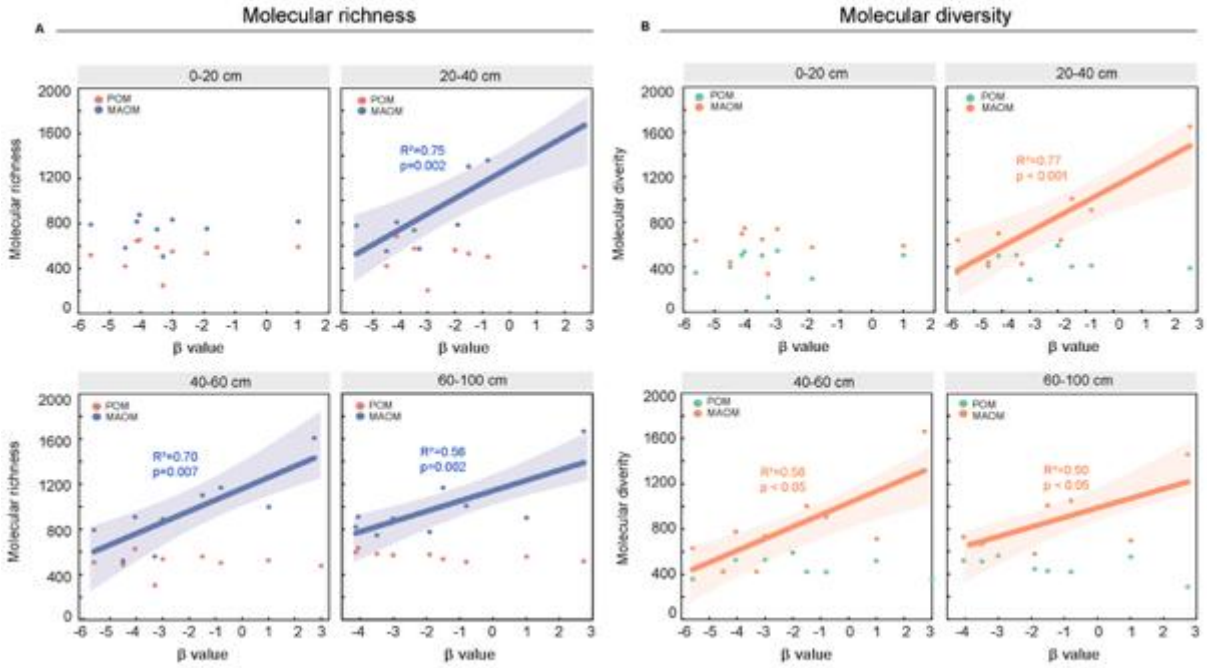
Supplementary Fig. 3. Molecular richness (A) and diversity (B) of POM and MAOM along the soil profiles in the grazing, 23-year, and 43-year exclosure treatment, respectively. With increasing soil depth carbon molecules became more distributed showing gradually increased molecular diversity. Organic compounds found in deeper soil horizons are typically order and more persistent than that in topsoil.



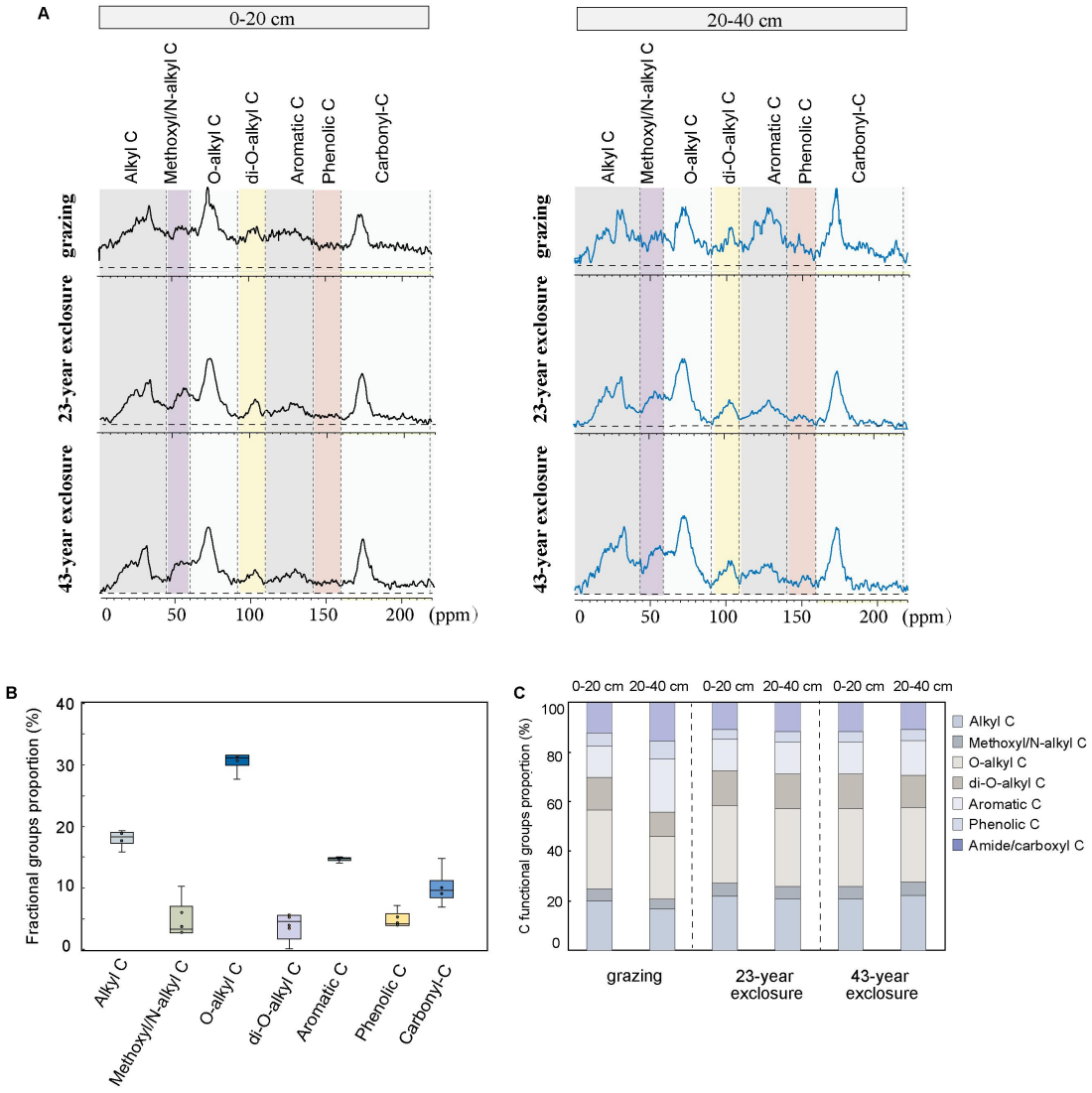
Supplementary Fig. 4. Investigation of the transformation of solid phase molecules during their passage to deep soil horizons by determining the corresponding shift of molecular mass distribution patterns in POM and MAOM. The molecular richness and diversity of POM and MAOM based on small molecular weight range of 0-5 kDa and large molecular weight range of 5-10 kDa distribution with soil depth in the grazing, 23-year, and 43-year exclosure treatment, respectively. The size of the circles represents the proportional scale. * indicates $p < 0.05$, **indicates $p < 0.01$, *** indicates $p < 0.001$.



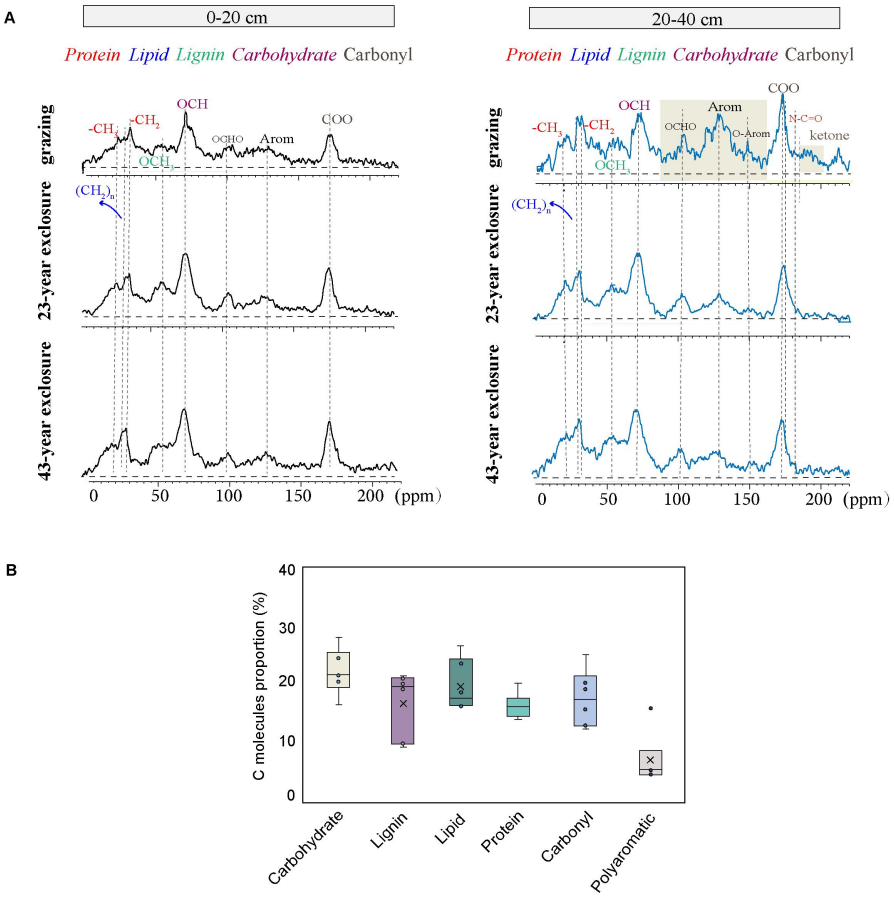
Supplementary Fig. 5. (A) The mass fractionation properties of SOM into light and heavy POM (LPOM and POM), as well as MAOM under grazing soil, 23-year and 43-year exclosure treatments, respectively, as well as variation across soil depths (0 - 20 cm, 20 - 40 cm, 40 - 60 cm, 60 - 100 cm). SOC fractionation percentage (B) of DOC, POC, and MAOC in the grazing soil, 23-year, and 43-year exclosure treatments along soil depth profile (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm), respectively.



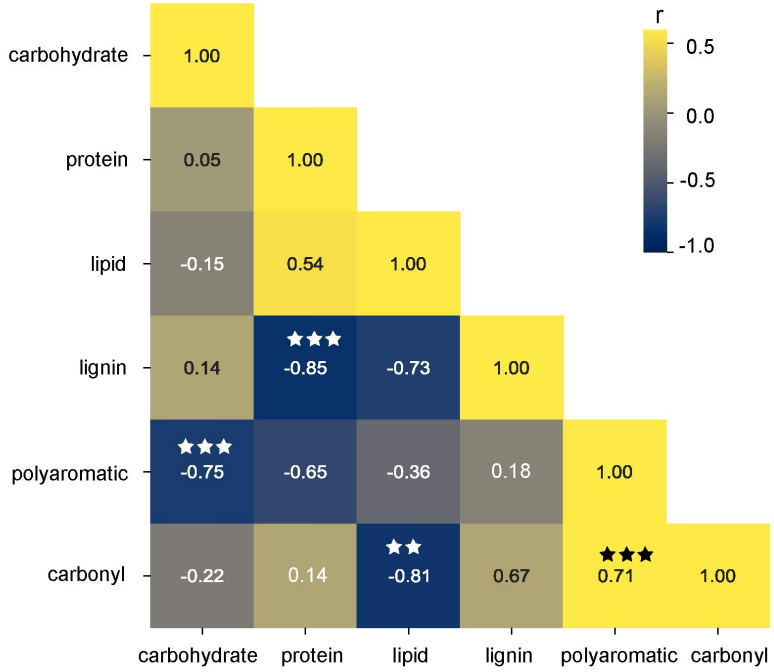
Supplementary Fig. 6. Correlation of β values of SOM with molecular richness (A) and molecular diversity (B) of POM and MAOM along the soil depth (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm), respectively. A lower β values indicate a faster decomposition rate of SOC, calculated by $\delta^{13}C$ divided by logarithmically converted SOC contents in five replicate sites for each soil layer under the corresponding treatments.



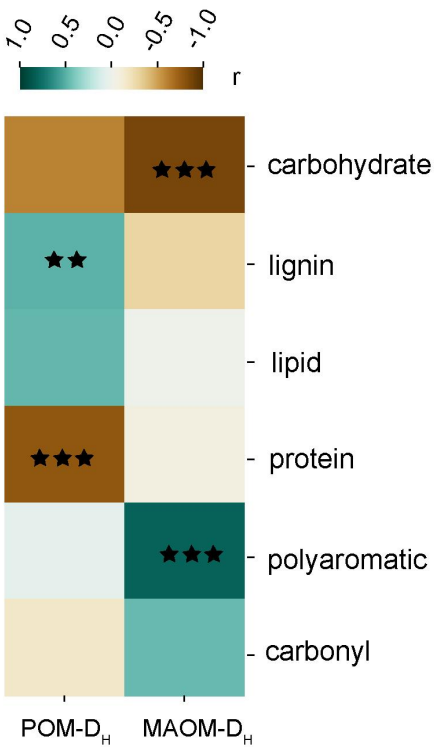
Supplementary Fig. 7. Solid-state ^{13}CNMR spectra displaying the chemical composition of soil samples across the grazing soil, 23-year exclosure, and 43-year exclosure treatments, along the soil depth profile (0 - 20 and 20 - 40 cm), respectively. After baseline correction, quantification was performed by dividing the spectra into seven chemical shift regions: 0 - 45, 45 - 60, 60 - 95, 95 - 110, 110 - 145, 145 - 165 and 165 - 215 ppm, assigned to alkyl C, N-alkyl + methoxyl C, O-alkyl C, Di-O-alkyl C, aromatic C, phenolic C and amide + carbonyl C, respectively.



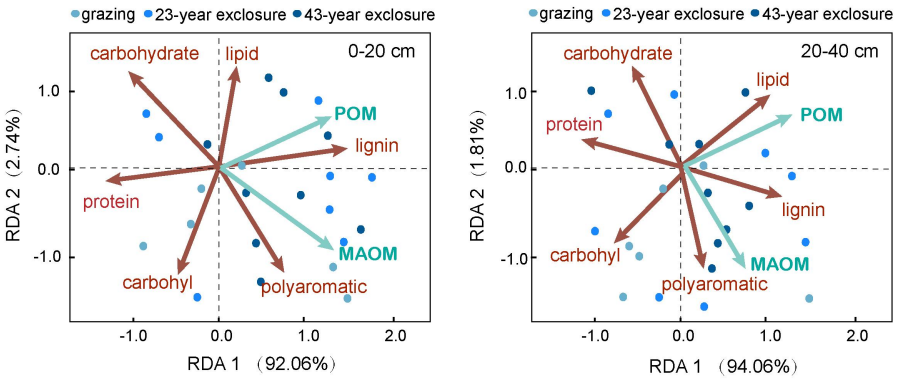
Supplementary Fig. 8. Molecules variations of organic matter revealed by quantitative solid-state ^{13}C NMR spectra within grassland soils and exclosure treatments. Carbon (C) compounds values by applying a molecular mixing model following the previous studies (Baldock et al., 2004; Fang et al., 2009; Hall et al., 2020). Displaying the carbon molecules of soil samples across the grazing soils and exclosure treatments, along the soil depth profile (0 - 20 and 20 - 40 cm), respectively. NMR Spectra mainly shows aromatic C at 130 ppm and carbonyl C at 172 ppm and over 60% of the aromatic C (between 145 and 100 ppm) is unprotonated, these aromatic products normally originate from protein and other polyphenols including mostly oxidized polyaromatic organic residues.



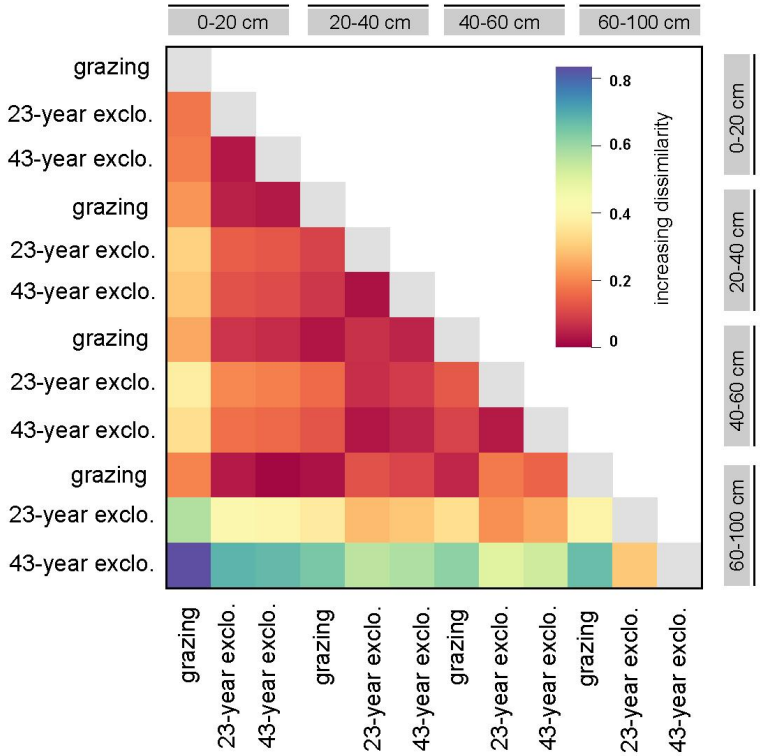
Supplementary Fig. 9. Correlation heatmap of each individual carbon molecules. The symbols ** and **** indicate corrected $p < 0.01$ and $p < 0.0001$, respectively.



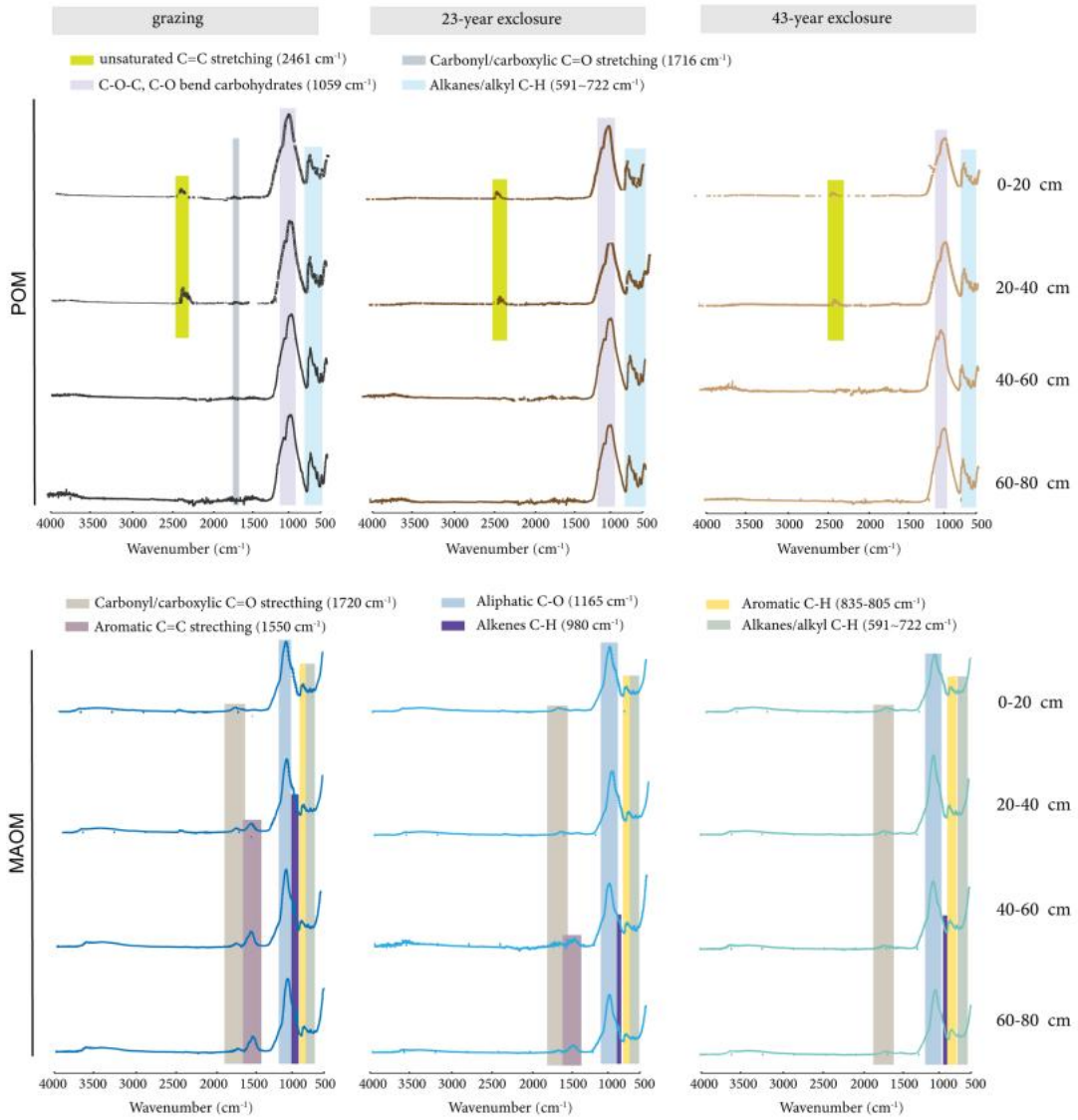
Supplementary Fig. 10. Correction relationship of individual carbon molecules and molecular diversity (D_H) of POM and MAOM. The symbols ** and *** indicated corrected $p < 0.05$ and $p < 0.001$, respectively.



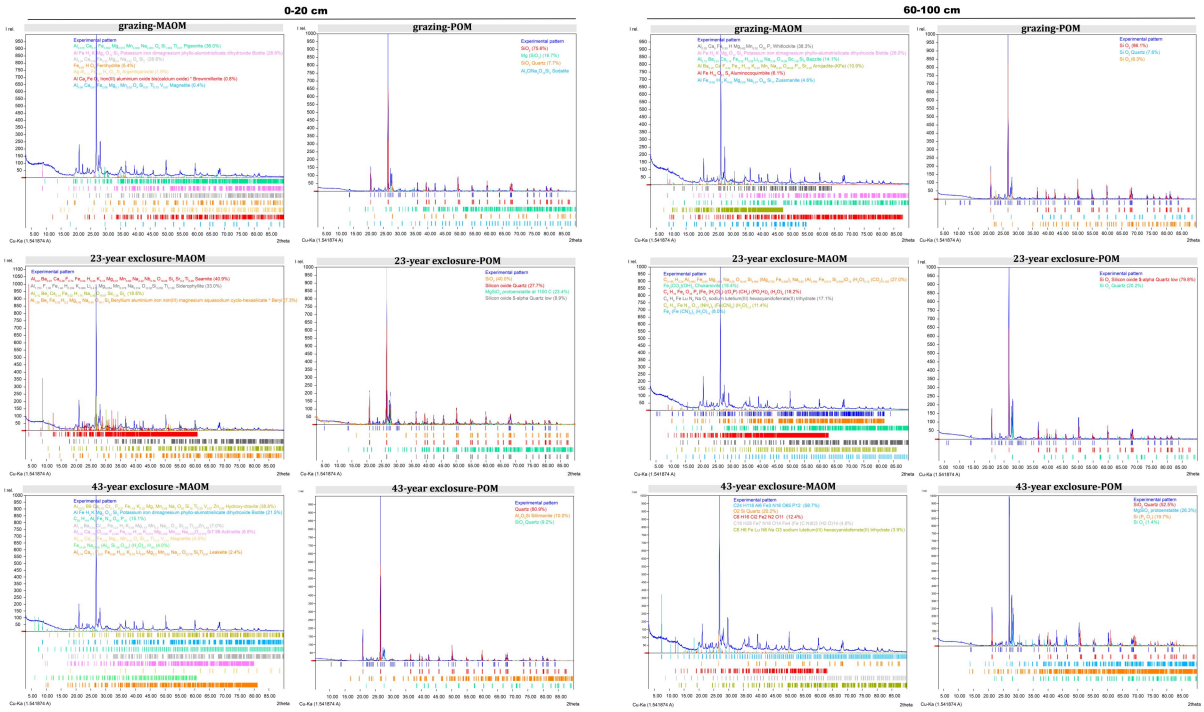
Supplementary Fig. 11. A significant correspondence between carbon divergences and the molecular diversity of POM and MAOM, respectively, along soil depth of 0 - 20 cm and 20 - 40 cm, by redundancy analysis.



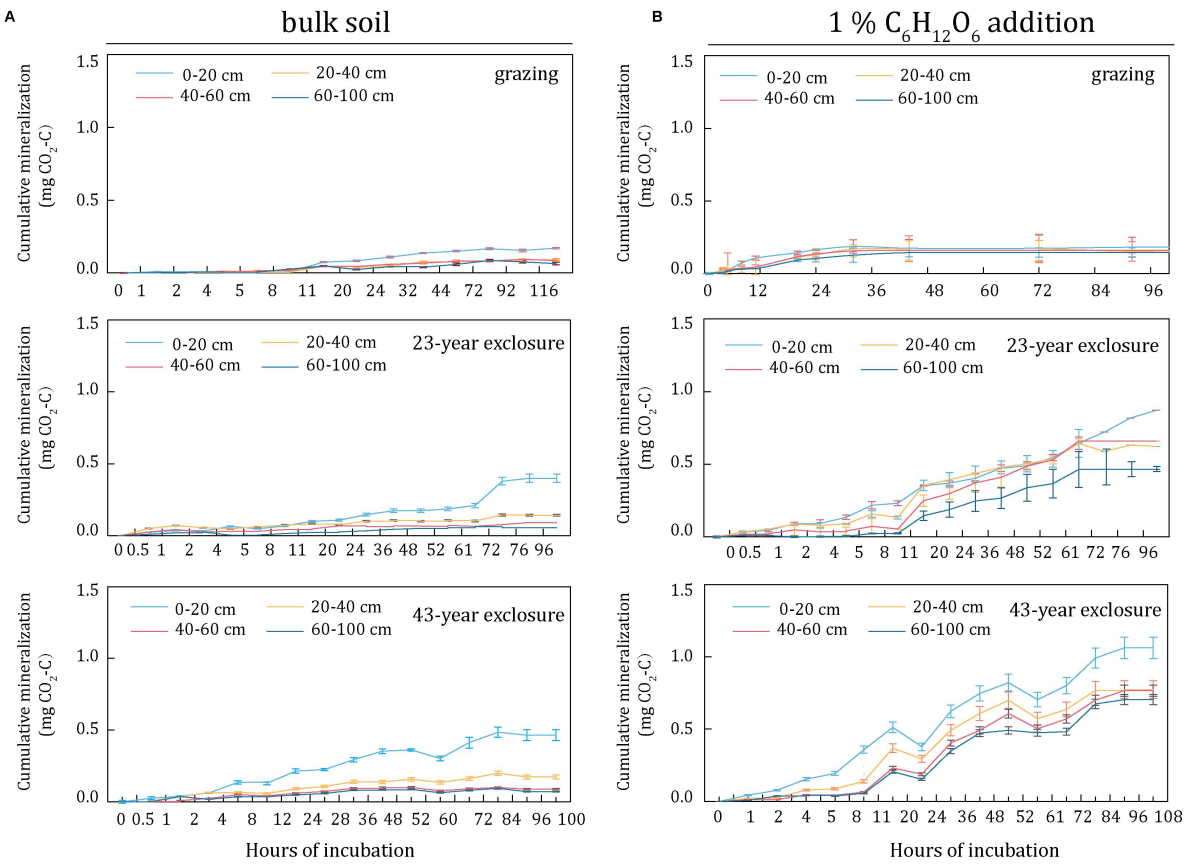
Supplementary Fig. 12. Distribution of SOM (based on molecular richness detected by MALDI-TOF-MS along the soil depths (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm) using Bray-Curtis Dissimilarity matrices.



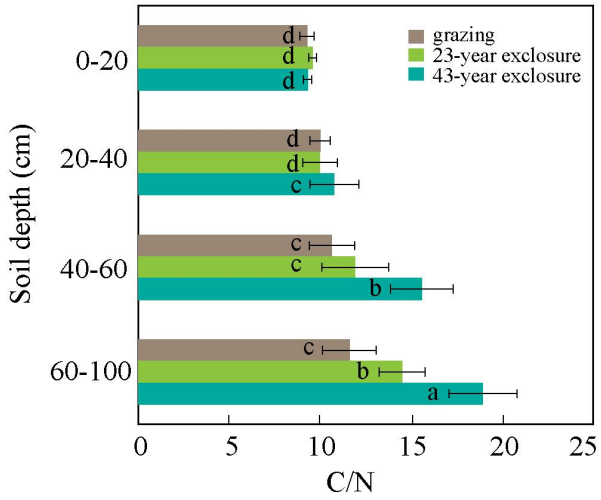
Supplementary Fig. 13. Fourier transform infrared attenuated total reflection (FTIR - ATR) spectra of POM and MAOM samples in the grazing soil, 23-year, and 43-year exclosure treatments along the whole soil depth (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm). Spectra are means of three replicate locations. Shared boxes indicate regions of interest, including unsaturated C=C stretching (~ 2361 cm^{-1}), carboxylic/carbonyl O=C-O bond (~ 1716 cm^{-1}, 1720 cm^{-1}), C-O bend carbohydrate (~ 1059 cm^{-1}), aliphatic C-O bonds (1165 cm^{-1}), aromatic C=C bonds (1590 - 1500 cm^{-1}) or amide C=O bonds (1660 - 1630 cm^{-1}), alkenes C-H bend region (980 cm^{-1}), aromatic C-H bonds (835 - 805 cm^{-1}), Si-O-Si bond (~ 800 cm^{-1}), alkanes/alkyl C-H bonds (591-722 cm^{-1}).



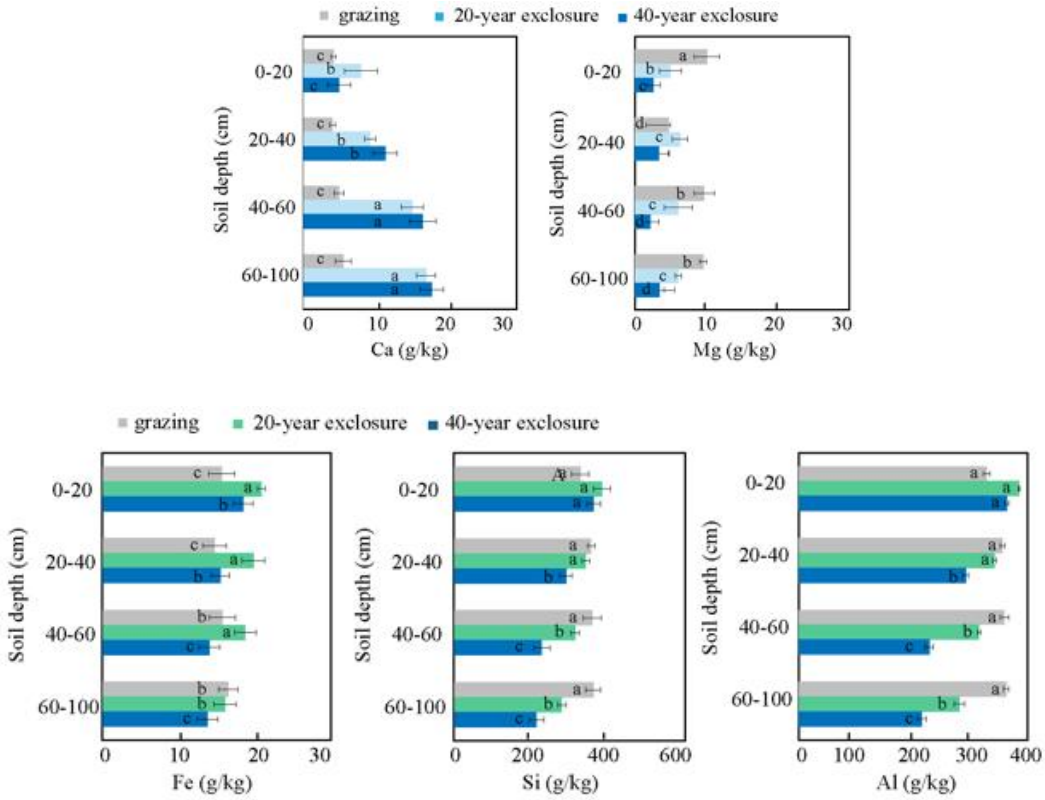
Supplementary Fig. 14. XRD analysis of POM and MAOM samples in the grazing soil, 23-year, and 43-year exclosure treatments, respectively.



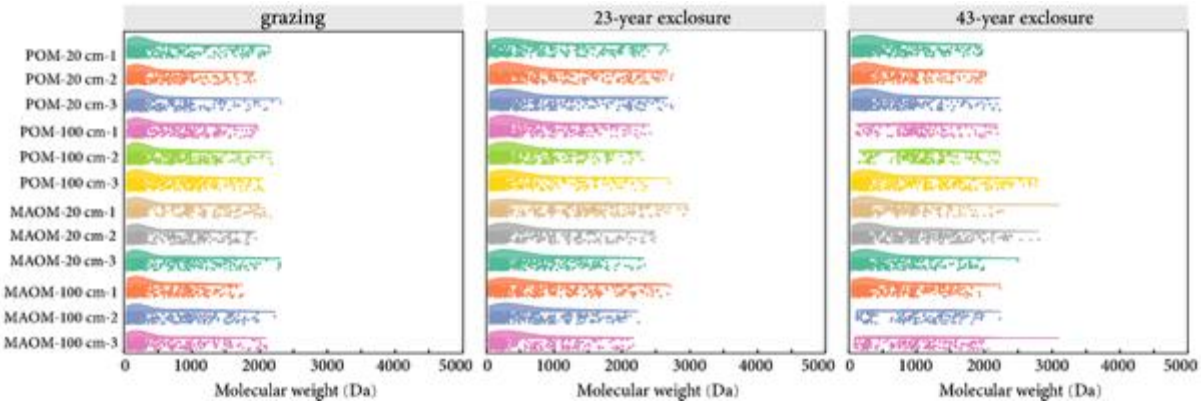
Supplementary Fig. 15. Cumulative mineralization (mg CO_2-C in 10 g soil) of bulk soil (A) and in presence of labile C (B) in the grazing soil, 23-year, and 43-year exclosure treatments, respectively. Glucose ($C_6H_{12}O_6$) was selected as labile C adding to 1% to bulk soil.



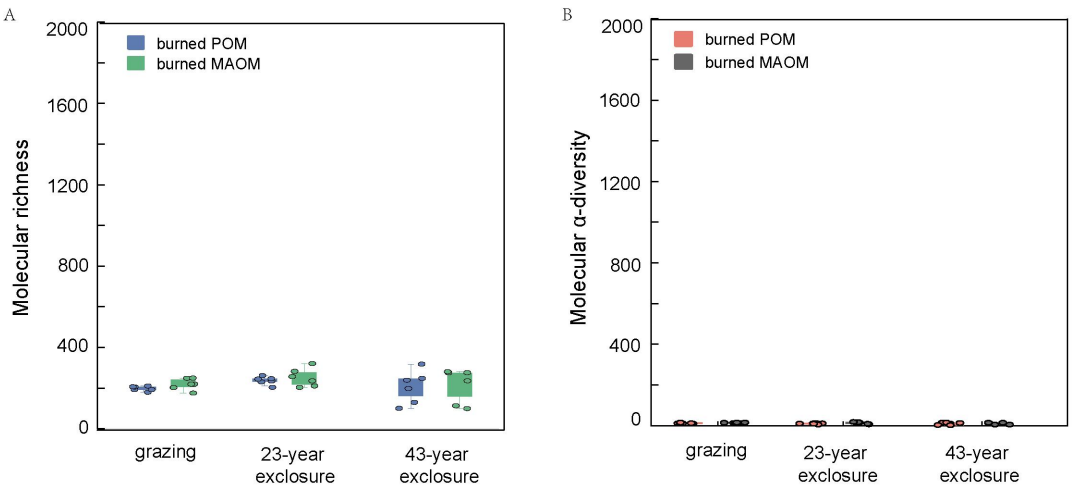
Supplementary Fig. 16. C/N ratios along the soil profiles (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm) in the grazing soil, 23-year, and 43-year exclosure treatments, respectively.



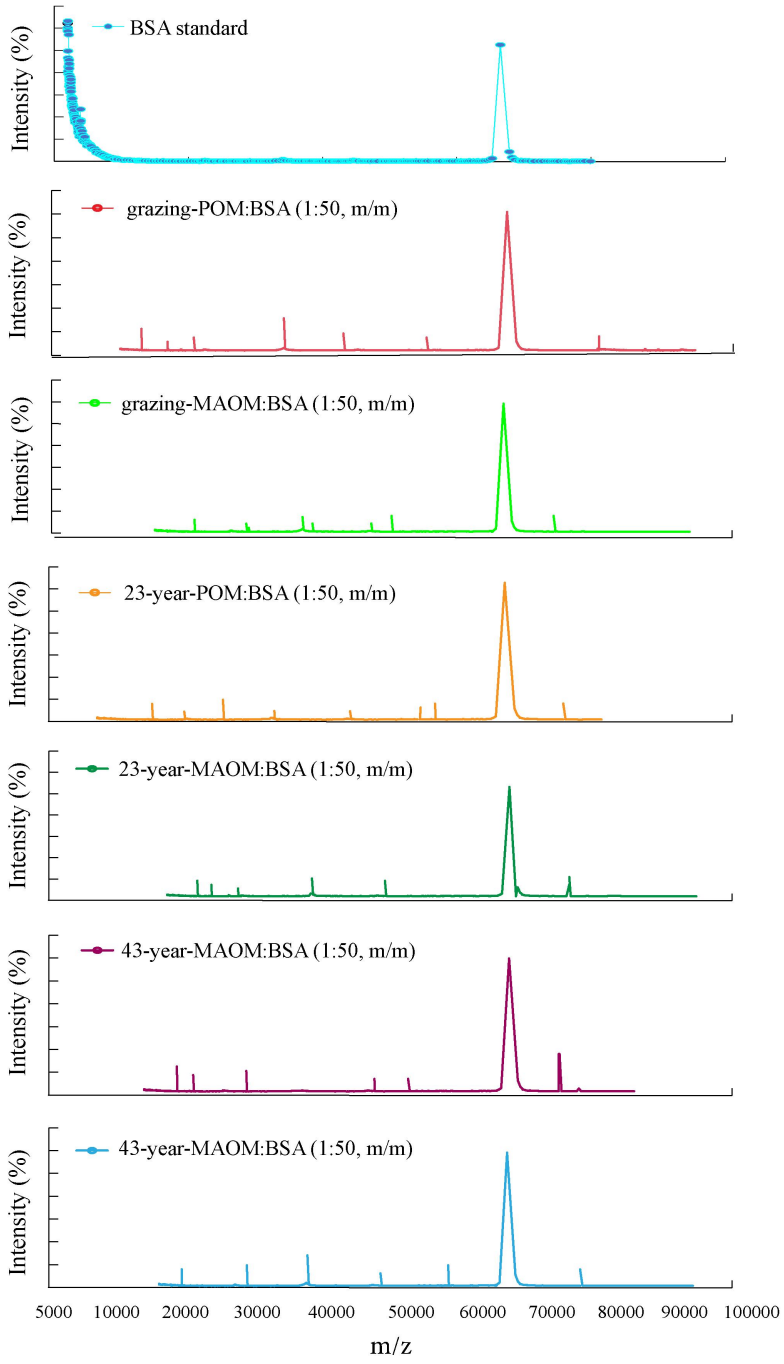
Supplementary Fig. 17. Total calcium (Ca), magnesium (Mg), iron (Fe), aluminum (Al), and silicon (Si) contents (g/kg soil) along the soil profiles (0 - 20, 20 - 40, 40 - 60, and 60 - 100 cm) in the grazing soil, 23-year, and 43-year exclosure treatments.



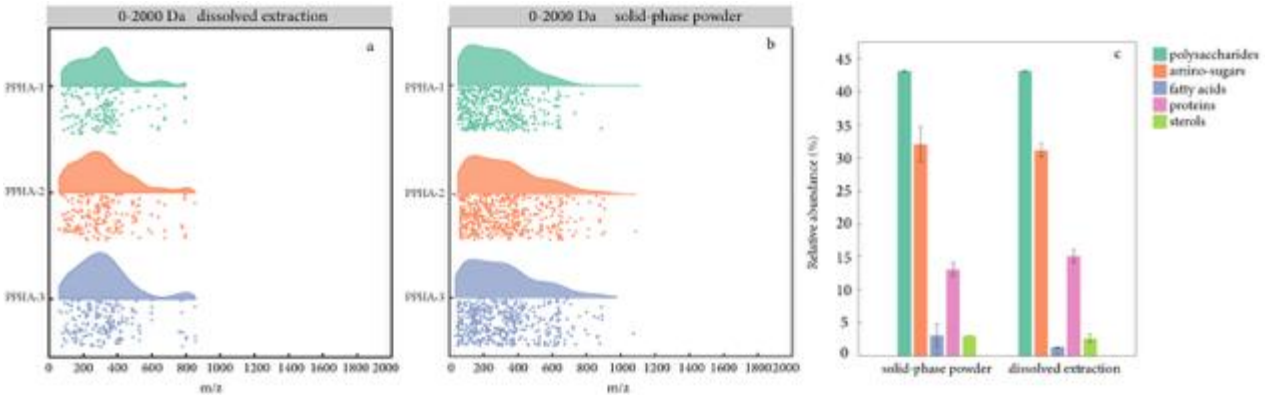
Supplementary Fig. 18. Molecular weight range (0 - 5 kDa) of burned samples. The spectrum is the identified background. No identified peak was detected in the 5-10 kDa of burned samples. Each POM or MAOM was burned at 900℃ to remove SOM. Each burned sample was determined following the same procedure as the identified background of samples.



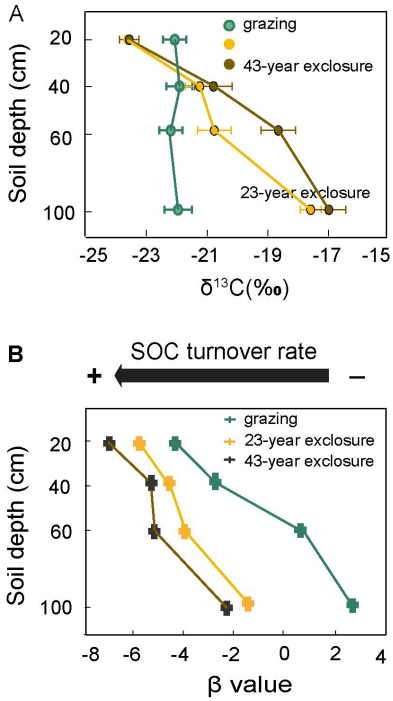
Supplementary Fig. 19. Molecular richness and diversity of burned POM and MAOM as the background references in the grazing, 23-year, and 43-year exclosure treatments, respectively. Each POM or MAOM was burned at 900°C to remove organic carbon. Each burned sample was determined following the same procedure as the identified background of samples. The scale of the Y-axis was set identical to that in supplementary Fig. 2 to facilitate the comparison of richness and diversity before and after POM and MAOM burning.



Supplementary Fig. 20. MALDI-TOF-MS spectra of bovine serium albumin (BSA, Sigma-Aldrich) standard and mixture samples of soil and BSA powder (1:50, m/m).



Supplementary Fig. 21. The IHSS Pahokee Peat Humic Acid (PPHA, standard 1S103H) was detected in the m/z range of 0-2 kDa for (a) the dissolved extraction following the previous method detected by MALDI-TOF-MS (Nicolau et al., 2015) and (b) The PPHA powder was fixed using carbon tape (NEM, FN731-5N)-coated with polymeric MALDI target plates and detected by MALDI-TOF-MS, respectively. (c) Comparison of relative abundance of soil composition between dissolved extraction and initial PPHA powder. Similar peaks were selected following the previous method (Nicolau et al., 2015).



Supplementary Fig. 22. (A) $\delta^{13}C$ and (B) β values along the soil depth under grazing soil, 23-year and 43-year exclosure treatments, respectively. The β value represents the carbon decomposition rate, calculated by dividing $\delta^{13}C$ by logarithmically converted SOC contents (Supplementary Methods 3-5), with measurements in five replicate sites within each soil layer. OM persistence during the exclosure process by examining the carbon turnover rate based on the variation of natural abundance of ^{13}C isotopes. A higher $\delta^{13}C$ in the exclosure treatments than that in the grazing treatment, indicating an increased SOC accumulation in the exclosure treatments. The $\delta^{13}C$ result was in line with a previous study conducted at similar field sites (Wang et al., 2017). The slope of linear regression (β value) between $\delta^{13}C$ values and organic carbon content was proposed as an indicator of carbon turnover in soils (Powers and Schlesinger, 2002; Wynn et al., 2005). Increased β value indicates a slower turnover rate, indicating an increased persistence of SOM (Garten et al.,2006).

Tables:

Supplementary Table S1. Ecology and molecular usage of diversity of POM and MAOM, followed by the same method published by Davenport et al. (2023).

Diversity index	Equation	Molecular context
α-diversity index Molecular richness	$D_R=S$	Counts known species or mass. Molecular identified peaks in MALDI-TOF MS spectra.
Molecular diversity Hill Numbers (D_{HN}, q=2)	${}^{q=2}D_{\mathrm{H}}(P)=(\sum_{i=1}^{S}P_i^q)^{1/(1-q)}$	Effective numbers of species or mass. Molecular richness and intensity are considered.
Bray-Curtis dissimilarity	$\mathrm{d}^{BCD}{}_{(\mathrm{i},j)}=\frac{\sum_{i=1}^{k}\lvert y_{i,k}-y_{j,k}\rvert}{\sum_{i=1}^{k}(y_{i,k}+y_{j,k})}$	Quantitative characteristics of species composition,(e.g.,abundance, coverage, and importance value), within the samples.

Supplementary Table S2. Corrected r values of predictor variables for molecular diversity indices separated by horizon for POM and MAOM from the MALDI-TOF-MS scan.

0-20 cm			
Predictors	**r (POM)**	**r (MAOM)**	**r (POM+MAOM)**
POC (mg C/ g soil)	0.78	0.87	1.00
MAOC (mg C/ g soil)	0.99	0.22	0.66
SOC ((mg C/ g soil)	1.00	0.36	0.76
pH	-0.60	0.53	0.06
WHC	0.96	0.60	0.91
C:N	0.81	-0.25	0.23
N:P	1.00	0.35	0.75
Fe (mg/g soil)	0.99	0.22	0.66
Al (mg/g soil)	**-0.94*****	0.65	**-0.93 *****
Ca (mg/g soil)	0.86	-0.16	0.32
Mg (mg/g soil)	**-0.80****	-0.85	**-0.80****

20-40 cm			
Predictors	**r (POM)**	**r (MAOM)**	**r (POM+MAOM)**
POC (mg C/ g soil)	0.99	0.93	0.96
MAOC (mg C/ g soil)	0.92	0.81	0.87
SOC ((mg C/ g soil)	0.87	0.73	0.80
pH	0.92	0.98	**0.96****
WHC	0.80	0.65	0.72
C:N	0.51	0.69	0.62
N:P	-0.99	-1.00	-1.00
Fe (mg/g soil)	0.59	**0.39****	**0.48****
Al (mg/g soil)	-0.99	-1.00	-1.00
Ca (mg/g soil)	0.97	1.00	0.99
Mg (mg/g soil)	0.73	0.56	0.64

40-60 cm			
Predictors	**r (POM)**	**r (MAOM)**	**r (POM+MAOM)**
POC (mg C/ g soil)	0.55	0.73	0.88
MAOC (mg C/ g soil)	0.94	-0.48	-0.23
SOC ((mg C/ g soil)	0.99	-0.31	-0.06
pH	-0.35	0.98	0.90
WHC	0.86	0.36	0.59
C:N	-0.16	1.00	0.97
N:P	0.02	-0.99	-0.99
Fe (mg/g soil)	0.89	**0.59****	**0.37****
Al (mg/g soil)	-0.44	-0.81	-0.94
Ca (mg/g soil)	0.50	0.78	0.91
Mg (mg/g soil)	-0.08	-0.97	-1.00

Table S2 (continued)

60-100 cm			
Predictors	**r (POM)**	**r (MAOM)**	**r (POM+MAOM)**
POC (mg C/ g soil)	-0.11	0.06	0.05
MAOC (mg C/ g soil)	-0.78	0.74	0.74
SOC ((mg C/ g soil)	-0.81	0.78	0.78
pH	-0.06	0.12	0.12
WHC	0.82	-0.85	-0.86
C:N	0.81	-0.77	-0.77
N:P	-0.82	0.90	0.92
Fe (mg/g soil)	-0.63	**0.59****	**0.58****
Al (mg/g soil)	-1.00	1.00	1.00
Ca (mg/g soil)	1.00	-0.99	-0.99
Mg (mg/g soil)	-0.91	0.88	0.88

References:

1. Baldock, J. A et al. Cycling and composition of organic matter in terrestrial and marine ecosystems. *Mar. Chem.* **92**, 39–64 (2004).
2. Davenport, R. et al. Decomposition decreases molecular diversity and ecosystem similarity of soil organic matter. *Proc. Natl. Acad. of Sci.* **120** (25), 11 (2023).
3. Fang, X. et al. Quantitative ^{13}C-NMR of whole and fractionated Iowa Mollisols for assessment of organic matter composition. *Geochim. Cosmochim. Acta* **74**, 584–598 (2010).
4. Hall, S.J. et al. Molecular trade-offs in soil organic carbon composition at continental scale. *Nat. Geosci.* **13**, 687–692 (2020).
5. Garten, Jr. C. T. Relationships among forest soil C isotopic composition, partitioning, and turnover times. *Can. J. For. Res.* **36**(9): 2157-2167.(2006).
6. Golubtsov,V.A. et al. Stable Carbon Isotopic Composition (δ^{13}C) as a Proxy of Organic Matter Dynamics in Soils on the Western Shore of Lake Baikal. *Eurasian Soil Sc.* **55**, 1700–1713 (2022).
7. Nicolau, R. et al. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) coupled to XAD fractionation: Method to algal organic matter characteriz ation. *Talanta* **136**, 102-107, 0039-9140 (2015).
8. Powers, J.S. & Schlesinger,W.H. Geographic and vertical patterns of stable carbon isotopes in tropical rain forest soils of Costa Rica. *Geoderma*, **109**: 141–160. (2002).
9. Wang, C. et al. Depth profiles of soil carbon isotopes along a semi-arid grassland transect in northern China. *Plant Soil* **417**, 43–52 (2017).
10. Wynn, J. G., Bird, M. I., & Wong, V. N. L. Rayleigh distillation and the depth profile of ^{13}C/^{12}C ratios of soil organic carbon from soils of disparate texture in iron range national park, far north Queensland, Australia. *Geochim. Cosmochim. Acta* **69**(8), 1961-1973 (2005).