Eco-Environment & Health 1 (2022) 212–218

Contents lists available at ScienceDirect

Eco-Environment & Health

journal homepage: www.journals.elsevier.com/eco-environment-and-health

Original Research Article

Molecular signature of soil organic matter under different land uses in the Lake Chaohu Basin

Han Gao [a,d], Huixin Li [a], Chen Lin [c], Pedro J.J. Alvarez [e], Caroline A. Masiello [f], Dongqiang Zhu [g], Ao Kong [b,**], Xiaolei Qu [a,*]

[a] State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, China
[b] School of Finance, Nanjing University of Finance and Economics, Nanjing 210023, China
[c] Key Laboratory of Watershed Geographic Sciences, Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China
[d] State Environmental Protection Key Laboratory of Soil Environmental Management and Pollution Control, Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China, Nanjing 210042, China
[e] Department of Civil and Environmental Engineering, Rice University, Houston TX 77005, USA
[f] Department of Earth, Environmental and Planetary Sciences, Rice University, Houston TX 77005, USA
[g] School of Urban and Environmental Sciences, Peking University, Beijing 100871, China

ARTICLE INFO

Keywords:
High-throughput untargeted mass spectrometry screening
Mass spectrum processing algorithm
Soil organic matter
Molecular signature
Land use

ABSTRACT

The concentration and molecular composition of soil organic matter (SOM) are important factors in mitigation against climate change as well as providing other ecosystem services. Our quantitative understanding of how land use influences SOM molecular composition and associated turnover dynamics is limited, which underscores the need for high-throughput analytical approaches and molecular marker signatures to clarify this etiology. Combining a high-throughput untargeted mass spectrometry screening and molecular markers, we show that forest, farmland and urban land uses result in distinct molecular signatures of SOM in the Lake Chaohu Basin. Molecular markers indicate that forest SOM has abundant carbon contents from vegetation and condensed organic carbon, leading to high soil organic carbon (SOC) concentration. Farmland SOM has moderate carbon contents from vegetation, and limited content of condensed organic carbon, with SOC significantly lower than that of forest soils. Urban SOM has high abundance of condensed organic carbon markers due to anthropogenic activities but relatively low in markers from vegetation. Consistently, urban soils have the highest black carbon/SOC ratio among these land uses. Overall, our results suggested that the molecular signature of SOM varies significantly with land use in the Lake Chaohu Basin, influencing carbon dynamics. Our strategy of molecular fingerprinting and marker discovery is expected to enlighten further research on SOM molecular signatures and cycling dynamics.

1. Introduction

Soil organic matter (SOM) is the largest carbon pool in the terrestrial environment, serving as an important carbon sink in the global carbon cycle. It also functions as a sink for hydrophobic organic pollutants, mitigating their bioavailability and environmental risks [1,2]. Land use activities have a significant influence on the SOM pool through changing plant diversity, soil erosion, and carbon turnover rate [3,4]. The impact of land use on the bulk contents of soil organic carbon (SOC) has been investigated [3,5–10]. The conversion of forest into farmland leads to considerable loss of SOC [3,5,6]. Current reports regarding the impact of urbanization on the SOC pool are not always consistent. Several studies reported higher SOC in urban soils than that in natural and agricultural lands [7,8], while others found similar [9] or lower [10] SOC in urban areas than non-urban areas. The molecular composition of SOM is one of the factors controlling its turnover time and accumulation in soils, and thus provides valuable insight into the SOC dynamics [11–13].

It is challenging to discern the influence of land use on the molecular composition of SOM due to the enormous heterogeneity of SOM, the analytical challenges in molecular characterization, as well as the complex and intertwined transformation mechanisms [11,14]. Advances in mass spectrometry provide an enormous opportunity to further our

* Corresponding author.
** Corresponding author.
E-mail addresses: aokong@nufe.edu.cn (A. Kong), xiaoleiqu@nju.edu.cn (X. Qu).

https://doi.org/10.1016/j.eehl.2022.10.003
Received 7 August 2022; Received in revised form 11 October 2022; Accepted 17 October 2022
Available online 17 November 2022

understanding of the evolution of SOM molecular composition [14]. Currently, high-resolution mass spectrometry and Fourier transform ion cyclotron resonance mass spectrometry are frequently used to determine the molecular composition of natural organic matter, but mostly focus on dissolved organic matter (DOM) [15–19]. The majority of the mass spectrometry methods focus on DOM other than SOM since the extraction process of SOM is often problematic due to the lower extraction efficiency and stronger bias [13]. Pyrolysis gas chromatography mass spectrometry (Py-GC-MS) is a fast, reproducible, and information-rich technique that can analyze the whole soil sample with limited pretreatment [20], which is suitable for molecular profiling of SOM. It was previously used to investigate the influence of land use on the molecular characteristics of SOM [21–27]. Generally, selected sets of pyrolysates were identified by the mass spectral libraries and grouped into classes (protein, lipids, phenolics, aromatics, lignin derivatives, etc.) to examine the origin and the evolution of SOM [20,24,28–30]. Another approach is to identify molecular markers in SOM which are the pyrolysates unique or prominent in a group of organic matter with known origin. It is in great need to further improve the throughput of SOM molecular profiling using Py-GC-MS and spectrum processing algorithm as well as discover molecular markers, with the aim of advancing quantitative understanding of SOM molecular signature (i.e., a small group of molecular markers strongly discriminating between SOM types) under the influence of land uses.

The Lake Chaohu Basin is located in the central part of Anhui Province, China, belonging to the drainage system of the Yangtze River. The lake itself is the fifth largest freshwater lake in China, with a surface area of 760 km^2 and a catchment area of 12,938 km^2 [31]. The Lake Chaohu Basin is a representative densely populated region in China, which processes a variety of land use patterns [32]. It accommodates a population of 9.65 million with a density more than 760 persons/km^2 and accounts for 24.65% of the gross domestic product in Anhui Province [32]. In recent decades, the land uses in the basin changed significantly owing to the expansion of agriculture, industry and urban areas [33]. Nevertheless, little is known regarding the impacts of land uses on the molecular signature and cycling dynamics of SOM in the Lake Chaohu Basin.

The objectives of this study were to (1) develop a high-throughput pipeline for profiling the molecular composition of SOM and molecular marker discovery, and (2) discern the molecular signatures of SOM in the Lake Chaohu Basin under different land uses. Seventy soil samples were collected from forests, farmlands, and urban areas distributed across the basin. We revealed the molecular fingerprints of SOM from different land uses and identified important molecular markers. The results show that land use drives changes of molecular composition of SOM, forming distinct patterns.

2. Materials and methods

2.1. Site description and sample collection

Seventy soil samples were collected across the Lake Chaohu Basin, including 25 forest sites, 24 farmland sites, and 21 urban sites (Fig. 1). The farmland sites were located at small farms growing wheat, soybean, corn, rice, and vegetables. The urban soils were collected from the residential areas and roadsides in cities and towns. Some of these sites were covered by bushes or grass, while others by sparse weeds. The dominant species at the forest sites were mainly pine, poplar, camphor, and China fir. The soils in the Lake Chaohu Basin are mainly Anthrosols rich in calcium and phosphate [34]. The majority of the tested samples were silt loam soils (i.e., 66 samples) except for two silty clay loam soils and two loam soils. Detailed site information and the soil texture analysis were summarized in Tables S1 – S2 and Fig. S1. For each site, nine soil samples (10 cm depth of the surface soil) were sampled by digging within a radius of 20 m and mixed into a single sample. Surface litter and vegetation were cleared prior to the sampling. The collected samples were air-dried

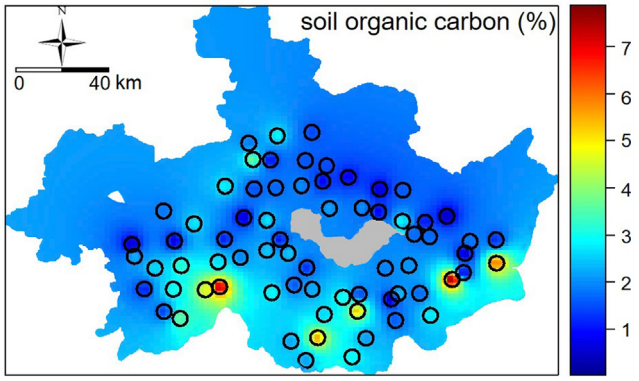

Fig. 1. Soil sampling sites and the corresponding soil organic carbon concentrations across the Lake Chaohu Basin. The grey area in the middle of the basin was the Lake Chaohu.

and ground with mortar and pestle into a fine powder. The SOC was measured using the hydrated heat potassium dichromate oxidation colorimetry method [35]. Soil pH was determined in a 1:2.5 soil : water suspension [36,37].

2.2. Py-GC-MS analysis

The Py-GC-MS based workflow for the molecular profiling and molecular marker discovery of SOM was summarized in Fig. 2. A Multi-shot Pyrolyser EGA/PY-3030D (Frontier Lab, Japan) coupled to an ISQ single quadrupole GC-MS system (Thermo Fisher Scientific, USA) was used for the Py-GC-MS analysis. The GC was fitted with a DB-5MS capillary column (0.25 mm × 30 m, 0.25 μm, J&K Scientific, USA) coupled to a quadrupole mass spectrometer. The helium carrier gas was used at a constant flow of 1.0 mL/min in the splitless mode. The GC temperature was programmed to hold at initial 35 °C for 5 min, increased at a rate of 2.5 °C/min to 200 °C, then increased at a rate of 5 °C/min to 270 °C and hold for another 5 min. The Py-GC-MS interface temperature, injector temperature, transfer line temperature, and source temperature were set at 280 °C, 250 °C, 280 °C, and 300 °C, respectively. The scan range of mass spectrum was from m/z 35 to 600 at 5 scans/s with an electron energy of 70 eV.

Soil samples (5.00 ± 0.10 mg) were weighed using a microbalance (XP56, Mettler Toledo, USA) and transferred into a deactivated stainless steel eco-cup (Eco-cup SF, Frontier Lab, Japan). Poly-α-methylstyrene (5 μL, 20 mg/L dispersed in dichloromethane, Mw 10,000, Sigma –Aldrich, USA) was injected into the eco-cup as an internal standard. A small amount of quartz wool was then inserted into the cup to prevent the sample from flying out. Samples were pyrolyzed at 610 °C for 12 s before the signal acquisition. The used eco-cups were burned using a butane high-temperature flame to clean. All peaks acquired by the Py-GC-MS were calibrated and normalized by the peak of Poly-α-methylstyrene.

2.3. Py-GC-MS spectrum processing algorithm

We developed a processing algorithm based on a previous study to achieve customized high-throughput spectrum processing (Fig. S2) [38]. The algorithm contains the following steps: 1) A normalization step enforces the constraint of equal total intensity for all spectra; 2) A smoothing step refers to a moving averaging procedure implemented for each spectrum to remove the high-frequency noise; 3) A baseline removal step is accomplished with the subtraction of a curve formed by the moving minima from each smoothed spectrum; 4) A peak detection step identifies strong peaks which have a significant signal to noise ratios (> 2) and peak slopes (> 0.01); 5) A peak alignment step adjusts the misalignment in retention times so that the detected peaks can be cross-referenced in

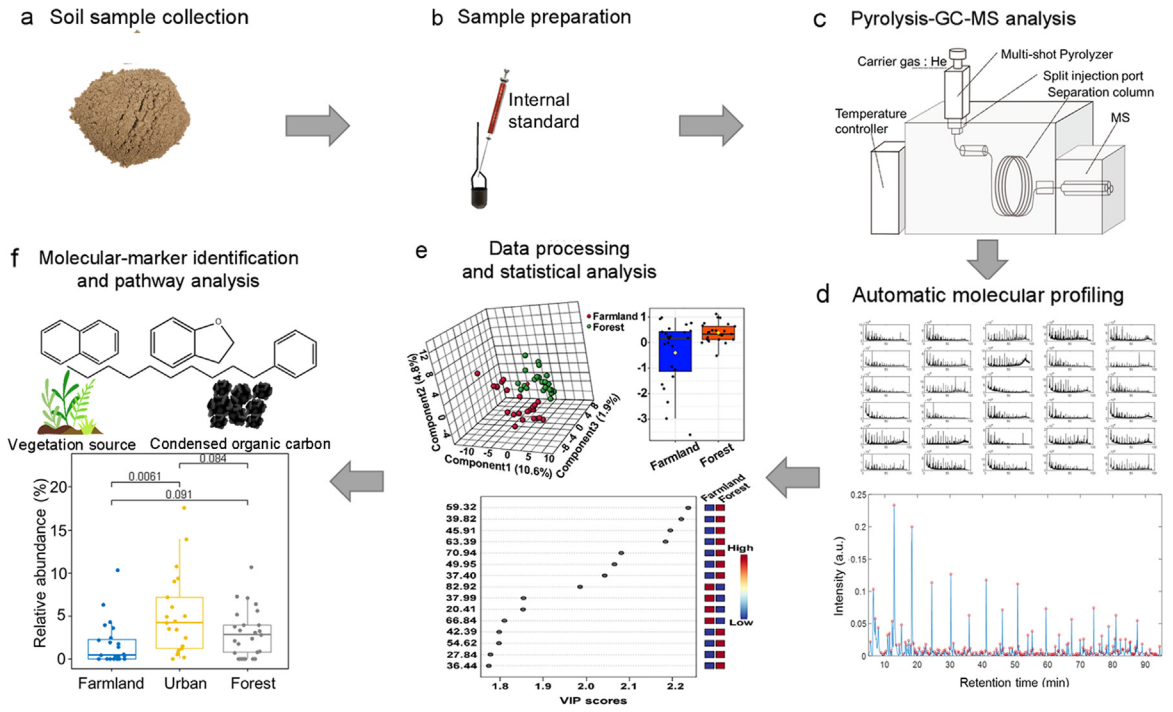

Fig. 2. Py-GC-MS based workflow for the molecular profiling and molecular marker discovery of soil organic matter. The workflow contains the following steps: a) The collected soil samples were air-dried and ground into a fine powder; b) Poly-α-methylstyrene was injected into the eco-cup as an internal standard; c) Py-GC-MS analysis; d) An in-house program based on Matlab was used to process the raw pyrograms and automatically extract detected peaks from the pyrograms to form a peak table; e) The peak table was analyzed by the MetaboAnalyst 4.0 and the relative abundance of molecules was compared among land uses; f) Pyrolysates were annotated using the NIST library, molecular markers were identified and related to potential soil organic matter origin and the relative abundance of the molecular marker was compared among land uses.

different samples; 6) At the end, all detected peaks from soil samples were automatically extracted and tabulated for further analysis. The algorithm code can be found in the Supplementary Data (SA). The compounds were annotated using the National Institute of Standards and Technology (NIST) mass spectral library with the best match.

2.4. Determination of soil black carbon (BC)

The BC concentrations in soil samples were determined using the thermochemical oxidation method (CTO-375) modified from previous studies [39,40]. Briefly, the inorganic carbonates in soils were first removed by acidifying 50 mg soil sample using 6 M HCl for 1 h at room temperature in a porcelain crucible with silica glaze surface (Purshee, China). The resulting soils were dried at 60 °C and then placed in a muffle furnace under 375 °C for 24 h with excess air. The resulting soil sample was analyzed using a CHN elemental analyzer (HT-1300/S-180403, Jena, Germany) at a combustion temperature of 950 °C to determine the BC content.

2.5. Statistical analysis

The tabulated peak list generated by the Py-GC-MS spectrum processing algorithm was analyzed by the MetaboAnalyst 4.0 (The Metabolomics Innovation Center, Alberta, Canada) [41]. The data were processed using interquartile range filtering, log-transformed, and auto-scaled. Partial least square discriminant analysis (PLS-DA) along with its R^2 and accuracy, as well as the variable importance in projection (VIP) were then performed on the MetaboAnalyst using the processed data. Student *t*-tests were used to determine if the means of data from two paired land uses were significantly different. The spatial statistical and geostatistical analysis was carried out using maptools [42] and gstat [43] packages in the *R* environment [44], respectively.

3. Results and discussion

3.1. SOC concentrations under different land uses

The geographical pattern of SOC across the Lake Chaohu Basin was summarized in Fig. 1. Forest soils contain 2.82% ± 1.95% SOC, which is significantly higher than farmland soils (1.86% ± 0.81%; $P = 0.047$; Fig. 3a). The difference between forest SOC (2.82% ± 1.95%) and urban SOC (2.00% ± 1.38%) was not statistically significant ($P > 0.05$). Soil pH may influence the biological activity and other processes that can affect SOM [45]. The soil pH values of the forest, farmland, and urban areas were 7.40 ± 0.84, 7.67 ± 0.51, and 8.00 ± 0.24, respectively (Fig. 3b). The pH was relatively high as the parent material in the Lake Chaohu Basin contains significantly amount of calcium carbonate [34], consistent with a previous report [46]. Forest soil pH was similar to farmland soil pH despite

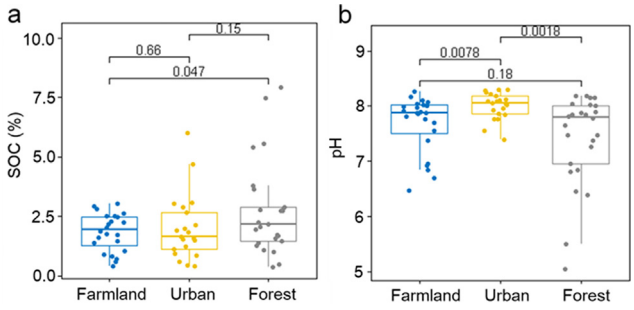

Fig. 3. The soil organic carbon (SOC) concentrations (a) and soil pH (b) under different land uses. The horizontal line in each box represents the median and the values on the top denote the *P* value of each two groups based on the *t*-test.

their different SOC concentrations (Fig. 3a). Urban soil pH was higher than that of farmland and forest soils ($P < 0.05$), most likely due to the leachates from calcareous materials [47]. The correlation between soil pH and SOC was poor with $R^2 = 0.2046$ (Fig. S3). Thus, soil pH was not the major determinant for the SOC pattern for different land uses.

3.2. Molecular fingerprints of SOM differed among land uses

The workflow for the molecular profiling and molecular marker discovery of SOM is illustrated in Fig. 2. The untargeted Py-GC-MS was used to obtain the raw pyrograms of SOM from different land uses (representative pyrograms for different land uses listed in Fig. S4). A Py-GC-MS spectrum processing algorithm was developed to process the raw pyrograms and automatically extract detected peaks from the pyrograms to form a molecular library in a high-throughput manner (Fig. S5). After processing all 70 SOM pyrograms, the algorithm identified 283 pyrolysates to form the molecular fingerprint for each SOM sample. Then, the program automatically formed a large molecular library of the SOM samples which contained 283 pyrolysates × 70 samples (Table S3). The molecular fingerprints of the SOM samples can be visualized by the heatmap as shown in Fig. S6.

Based on the molecular structure, the SOM fingerprints were grouped into twelve classes. The relative abundances (i.e., the peak intensity divided by the sum of all peak intensities in the pyrogram) of these classes were summarized in Fig. 4a. The fingerprints of SOM can provide some insights on the molecular patterns of different land uses. Among these twelve classes, the relative abundances of five classes of pyrolysates were statistically different in paired land use groups, including alcohols, furans, phenols, benzenes, and ketones (Fig. 4b–f). Furans and phenols are typical pyrolysis products from polysaccharides and lignin, respectively [48]. Alcohols are typical pyrolysis products from holocellulose and the decarboxylation products of polycarboxylic acids [49,50]. Ketones are often found in the pyrogram of manure, crop residue, and oil shale [51–53]. Benzenes are a group of common pyrolysates widely found for SOM with different origins. Farmland SOM has significantly lower alcohols and phenols than urban and forest SOM ($P < 0.05$, Fig. 4b–d). This result suggests that farmland SOM has lower input from holocellulose and lignin originating from wood materials than urban and forest SOM. Forest SOM contains higher furans but lower benzenes than farmland SOM ($P < 0.05$, Fig. 4c–e). Thus, forest SOM is expected to contain significantly higher polysaccharides than farmland SOM. Farmland SOM fingerprints have the highest ketones among all land uses ($P < 0.05$, Fig. 4f), most likely due to the higher input from crop residue.

3.3. Different land uses result in specific molecular signatures of SOM

The fingerprints of SOM samples were then analyzed by PLS-DA. SOM from different land uses can be separated from each other by PLS-DA with decent accuracy (Fig. 5). The PLS-DA accuracy of urban and forest soils is relatively low due to the presence of three urban soil samples around the classification boundary. The small overlap between shaded ellipses for each paired land uses suggests that land use has a significant influence on the molecular composition of SOM. Among the 283 pyrolysates in the SOM molecular fingerprints, 47 pyrolysates contributed significantly to the separation of SOM from different land uses (VIP score >1.7 for a paired land uses, $P < 0.05$; a full list of these pyrolysates can be found in Table S4). Here, the VIP scores were computed on the first PLS-DA component because it always dominates the explanatory power in all paired cases. Nevertheless, it is challenging to relate these pyrolysates to the specific origins of SOM. Among them, we are able to identify five molecular markers that can be associated with structural precursors representing different origins according to the literature [54–58] (Table 1). We grouped the markers to represent two origins affecting the structure of SOM, which were vegetation production and condensed organic carbon (e.g., black carbon, kerogen, coal, and soot). The relative abundances of markers were compared among different land uses as shown in Fig. 6.

The vegetation production marker group includes naphthalene and 2,3-dihydrobenzofuran. High yield of naphthalene is commonly found in the pyrolysis of lignin, indicating vegetation production [58]. Forest SOM fingerprints had higher naphthalene than urban SOM ($P = 0.024$, Fig. 6a). 2,3-dihydrobenzofuran is a polysaccharide-derived pyrolysate related to the vegetation production [59,60]. Its relative abundance follows the order of forest SOM, farmland SOM, urban SOM ($P < 0.05$, Fig. 6b). The condensed organic carbon marker group includes 1-methoxy-13-methyl-pentadecane and alkylbeneznes (i.e.,

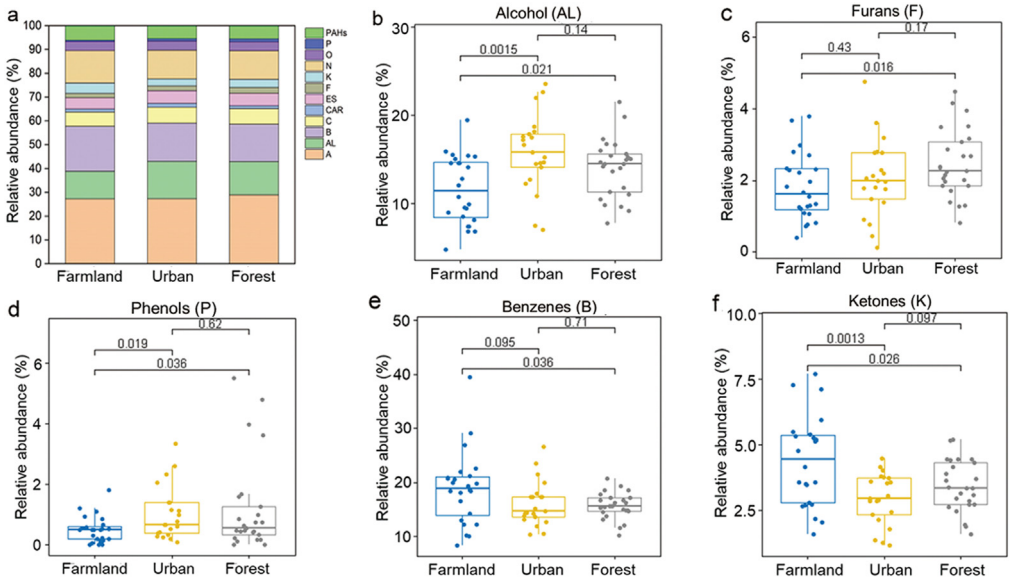

Fig. 4. Relative abundance of (a) all pyrolysate classes, (b) alcohol, (c) furans, (d) phenols, (e) benzenes, (f) ketones in soil organic matter under different land uses. The horizontal line in each box represents the median and the values on the top of the figure denote the *P* value of each two groups based on the *t*-test. PAHs, polycyclic aromatic hydrocarbons; P, phenols; O, other compounds; N, nitrogenous compounds; K, ketones; F, furans; ES, esters; CAR, carboxylic compounds; C, alicyclic compounds; B, benzenes; AL, alcohol; A, aliphatic compounds.

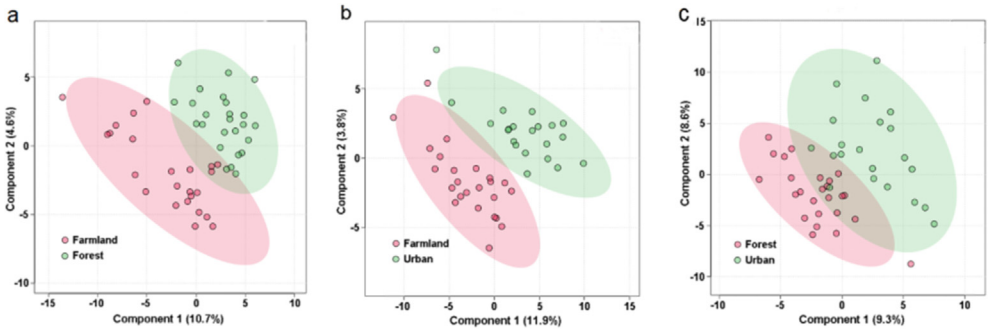

Fig. 5. Partial least square discriminant analysis (PLS-DA) score plots of (a) farmland and forest soils (accuracy = 0.58, R^2 = 0.78), (b) farmland and urban soils (accuracy = 0.72, R^2 = 0.83), and (c) urban and forest soils (accuracy = 0.40, R^2 = 0.70). Shaded ellipses represent a 95% confidence interval for each type of land use. The performance achieved with two axis was reported for each paired land use.

Table 1
Molecular markers identified for differentiating SOM molecular composition in forest, farmland and urban soils.

Origin	Retention time (min)	Molecular markers	VIP score	Class
Vegetation	30.60	Naphthalene	2.1109	PAHs
	25.98	2,3-dihydrobenzofuran	1.7106	F
Condensed organic carbon	69.75	1-methoxy-13-methyl-pentadecane	2.4854	A
	39.82	Heptylbenzene	2.2528	B
	54.62	Decylbenzene	1.7130	B

VIP, variable importance in projection; PAHs, polycyclic aromatic hydrocarbons; F, furans; A, aliphatic compounds; B, benzenes.

heptylbenzene and decylbenzene). 1-methoxy-13-methyl-pentadecane is related to fossil fuel combustion [54] and fire events [59]. It is more abundant in urban SOM fingerprints than those in farmland (P = 0.0061, Fig. 6c). Alkylbeneznes are rich in the pyrograms of kerogens and coals [55–57]. The relative abundance of heptylbenzene followed the order of urban SOM, forest SOM, farmland SOM (Fig. 6d). Forest SOM had significantly more decylbenzene than farmland SOM (P = 0.0098, Fig. 6e). SOM influenced by different land uses has distinct signatures. Forest SOM has high abundance of the markers for vegetation production, including naphthalene and 2,3-dihydrobenzofuran. The contribution of plant-derived SOM was lower in farmlands and urban areas than that in forests, consistent with a previous report [61]. Urban SOM generally have abundant markers for condensed organic carbon, especially that from fossil fuel combustion (i.e., 1-methoxy-13-methyl-pentadecane).

We further explored the potential accumulation mechanisms of SOM under different land uses, which lead to their distinct molecular signatures. Land use is the most important driver of changes in biodiversity [62]. In the Lake Chaohu Basin, forests usually have the highest plant diversity, resulting in abundant and diverse vegetation production, contributing to the molecular pool of SOM [3]. The plant diversity of these farmland and urban areas was relatively low as compared with that of forests due to long-term monoculture and urban development (Table S1), respectively. This is consistent with the high vegetation production markers in forest SOM. Additionally, wildfires generate condensed organic carbon in forests such as black carbon which is fairly difficult to degrade [63,64]. Thus, forest also had significant amount of markers for condensed organic carbon (i.e., heptylbenzene and decylbenzene). The diverse and rich carbon inputs of both fast-cycling and decomposition-resistant carbon in forests result in the highest SOC concentration (Fig. 3a). The functional complexity of the forest SOM

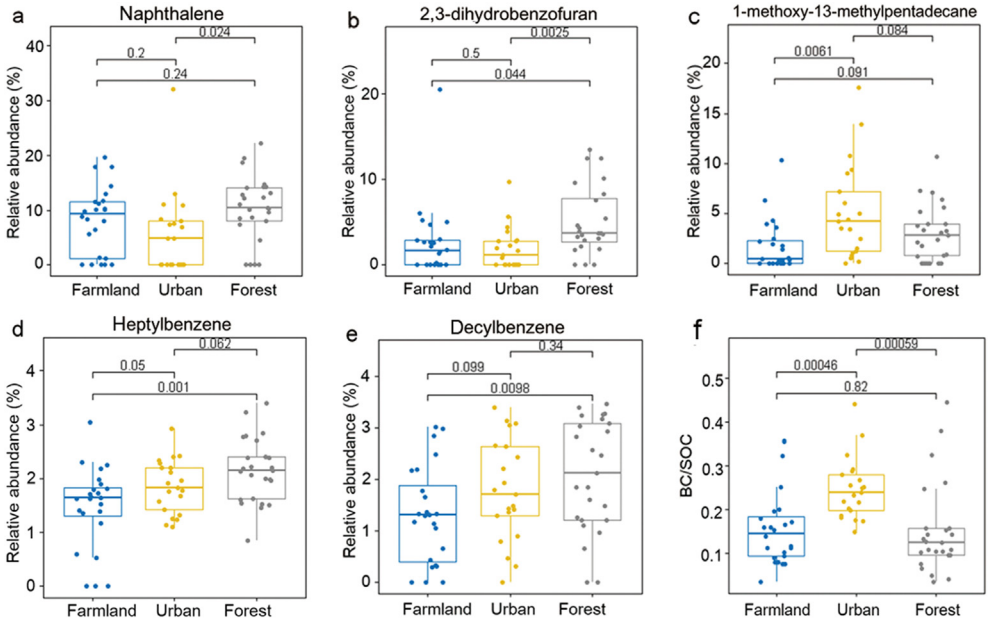

Fig. 6. Relative abundance of the molecular markers in soil organic matter under different land uses: (a) naphthalene, (b) 2, 3-dihydrobenzofuran, (c) 1-methoxy-13-methyl-pentadecane, (d) heptylbenzene, (e) decylbenzene, and (f) the black carbon/soil organic carbon (BC/SOC) ratios of soil organic matter under different land uses. The horizontal line in each box represents the median, and the values on the top of the figure denote the P value of each two groups based on the t-test.

may also contribute to its persistence and facilitates the accumulation of SOC [65].

Urban SOM had abundant markers for condensed organic carbon. The soil BC contents were determined to further examine the potential role of condensed organic carbon in SOM under different land uses. There is a significant correlation between BC/SOC values and the relative abundance of 1-methoxy-13-methyl-pentadecane in SOM pyrograms ($P = 0.0001$, Fig. S7), although the R^2 is relatively low. This result further supported the role of 1-methoxy-13-methyl-pentadecane as a molecular marker for condensed organic carbon. The BC/SOC ratios in urban, forest, and farmland soils were 0.25 ± 0.07, 0.15 ± 0.10, and 0.16 ± 0.09, respectively. Urban SOM had significantly higher BC/SOC ratio than forest and farmland SOM ($P < 0.05$, Fig. 6f), generally consistent with the results of molecular markers. It agrees with previous studies reporting that condensed organic carbon accumulates in industrial zones, residential sites, and roadsides in urban areas [8,66]. The BC/SOC ratio in urban SOM in the Lake Chaohu Basin is close to that in urban soils in Beijing (0.31–0.39) [67] and UK cities (0.28–0.39 for cities across the North East of England) [68]. The major source of condensed organic carbon in urban areas is wood and fossil fuel combustion [8]. Thus, the signature of urban SOM is closely related to anthropogenic activities, leading to the accumulation of decomposition-resistant carbon in urban soils [68].

The ratio of BC/SOC was similar between farmland and forest soils (Fig. 6f). The major source of soil BC in farmland is crop residue burning [69], which was a previously used agricultural practice in the Lake Chaohu Basin. The BC/SOC ratio in an agricultural region in central France was reported to be from 0.01 to 0.32 with an average value of 0.05 [69]. The BC/SOC ratio in farmland SOM in the Lake Chaohu Basin was higher than that in central France, most likely due to more frequent crop residue burning. Condensed organic carbon generated during crop residue burning is less stable than that generated during fossil fuel combustion in urban areas. Our previous studies suggested that condensed organic carbon generated during biomass combustion will gradually release dissolved black carbon which was readily subject to photodegradation and erosion [70,71]. Farmland soils were expected to have more vegetation input than the urban soils as more than half of the urban soil sites in this study only had sparse weeds (Table S1). Urban soils had more abundant condensed organic carbon, leading to similar SOC concentrations in these two land uses.

4. Conclusions

In the present study, we examined the molecular signatures of SOM under forest, farmland, and urban land uses in the Lake Chaohu Basin.

(1) We first developed a customized spectrum processing algorithm to process the raw pyrograms of SOM samples and extract detected peaks from the pyrograms in a high-throughput manner. The algorithm identified 283 pyrolysates to form the molecular fingerprints for each SOM sample and automatically formed a large molecular library (283 pyrolysates × 70 samples).

(2) SOM fingerprints from different land uses had specific patterns. Farmland SOM fingerprints had the lowest alcohol/phenols and the highest ketones among different land uses. Forest SOM fingerprints contained higher furans but lower benzenes than farmland SOM. Farmland SOM has lower input from holocellulose and lignin originating from wood materials but higher input from crop residue than urban and forest SOM.

(3) SOM fingerprints from different land uses can be clearly separated from each other by PLS-DA analysis. Five pyrolysates were identified as molecular markers that can be used to distinguish the relative importance of specific SOM sources. There is a significant correlation between BC/SOC values and the relative abundance of one of the markers for condensed organic carbon, i.e., 1-methoxy-13-methyl-pentadecane.

(4) Forest SOM had abundant markers from both vegetation and condensed organic carbon, leading to its highest SOC content. Urban SOM had high abundance of condensed organic carbon markers due to anthropogenic activities but relatively low in markers from vegetation. Consistently, urban soils have the highest black carbon/SOC ratio.

Our results highlight the potential impacts of land uses on the SOM signature. Future study is needed to further explore the relationships between the discerned markers and specific carbon sources, which help understand the complex SOM responses to land uses. Overall, the molecular fingerprinting and marker discovery strategy presented here can enlighten further research on how large-scale land use planning projects affect SOM molecular signature and cycling dynamics, and inform the potential for broader inclusion of land use planning in greenhouse gas policies.

Author contributions

HG: Investigation, Methodology, Writing—original draft. HL: Investigation. CL: Investigation. PJA: Writing—review & editing. CAM: Writing—review & editing. DZ: Writing—review & editing. AK: Conceptualization, Investigation, Funding acquisition, Software, Writing—review & editing. XQ: Conceptualization, Methodology, Supervision, Funding acquisition, Writing—review & editing.

Declaration of competing interests

Authors declared no conflict of interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (grant nos. 2019YFC1804201, 2020YFC1807002), China Postdoctoral Science Foundation (grant no. 2021M701670), the National Natural Science Foundation of China (grant no. 21876075), and Jiangsu Planned Projects for Postdoctoral Research Funds (grant no. 2021K357C).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eehl.2022.10.003.

References

[1] D.E. Kile, C.T. Chiou, H.D. Zhou, H. Li, O.Y. Xu, Partition of nonpolar organic pollutants from water to soil and sediment organic matters, Environ. Sci. Technol. 29 (1995) 1401–1406.
[2] Y.P. Chin, C.S. Peven, W.J. Weber, Estimating soil/sediment partition coefficients for organic compounds by high performance reverse phase liquid chromatography, Water Res. 22 (1988) 873–881.
[3] M. Lange, N. Eisenhauer, C.A. Sierra, H. Bessler, C. Engels, R.I. Griffiths, P.G. Mellado-Vazquez, A.A. Malik, et al., Plant diversity increases soil microbial activity and soil carbon storage, Nat. Commun. 6 (2015) 6707.
[4] W.M. Post, K.C. Kwon, Soil carbon sequestration and land-use change: processes and potential, Global Change Biol. 6 (2000) 317–327.
[5] A. Don, J. Schumacher, A. Freibauer, Impact of tropical land-use change on soil organic carbon stocks-a meta-analysis, Global Change Biol. 17 (2011) 1658–1670.
[6] J.A. Foley, R. DeFries, G.P. Asner, C. Barford, G. Bonan, S.R. Carpenter, F.S. Chapin, M.T. Coe, et al., Global consequences of land use, Science 309 (2005) 570–574.
[7] V.I. Vasenev, J.J. Stoorvogel, I.I. Vasenev, Urban soil organic carbon and its spatial heterogeneity in comparison with natural and agricultural areas in the Moscow region, Catena 107 (2013) 96–102.
[8] V. Vasenev, Y. Kuzyakov, Urban soils as hot spots of anthropogenic carbon accumulation: review of stocks, mechanisms and driving factors, Land Degrad. Dev. 29 (2018) 1607–1622.
[9] L.F. Weissert, J.A. Salmond, L. Schwendenmann, Variability of soil organic carbon stocks and soil CO2 efflux across urban land use and soil cover types, Geoderma 271 (2016) 80–90.
[10] H.K. Jo, Impacts of urban greenspace on offsetting carbon emissions for middle Korea, J. Environ. Manag. 64 (2002) 115–126.
[11] M.J. Simpson, A.J. Simpson, The chemical ecology of soil organic matter molecular constituents, J. Chem. Ecol. 38 (2012) 768–784.
[12] M.W.I. Schmidt, M.S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I.A. Janssens, M. Kleber, I. Kögel-Knabner, et al., Persistence of soil organic matter as an ecosystem property, Nature 478 (2011) 49.
[13] J. Lehmann, M. Kleber, The contentious nature of soil organic matter, Nature 528 (2015) 60.

[14] K. Mopper, A. Stubbins, J.D. Ritchie, H.M. Bialk, P.G. Hatcher, Advanced instrumental approaches for characterization of marine dissolved organic matter: extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy, Chem. Rev. 107 (2007) 419–442.
[15] J. D'Andrilli, C.M. Foreman, A.G. Marshall, D.M. McKnight, Characterization of IHSS Pony Lake fulvic acid dissolved organic matter by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and fluorescence spectroscopy, Org. Geochem. 65 (2013) 19–28.
[16] L.M. Lynch, N.A. Sutfin, T.S. Fegel, C.M. Boot, T.P. Covino, M.D. Wallenstein, River channel connectivity shifts metabolite composition and dissolved organic matter chemistry, Nat. Commun. 10 (2019) 459.
[17] 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 (2014) 3804.
[18] M. Gonsior, J. Valle, P. Schmitt-Kopplin, N. Hertkorn, D. Bastviken, J. Luek, M. Harir, W. Bastos, et al., Chemodiversity of dissolved organic matter in the Amazon Basin, Biogeosciences 13 (2016) 4279–4290.
[19] X.M. Li, G.X. Sun, S.C. Chen, Z. Fang, H.Y. Yuan, Q. Shi, Y.G. Zhu, Molecular chemodiversity of dissolved organic matter in paddy soils, Environ. Sci. Technol. 52 (2018) 963–971.
[20] S. Derenne, K. Quénéa, Analytical pyrolysis as a tool to probe soil organic matter, J. Anal. Appl. Pyrolysis 111 (2015) 108–120.
[21] Z.S. Zhang, J.J. Wang, X.G. Yu, M. Jiang, J. Bhadha, A. Wright, Impacts of land use change on soil organic matter chemistry in the Everglades, Florida-a characterization with pyrolysis-gas chromatography mass spectrometry, Geoderma 338 (2019) 393–400.
[22] C.P. de Assis, J.A. Gonzalez-Perez, J.M. de la Rosa, I. Jucksch, E.D. Mendonca, F.J. Gonzalez-Vila, Analytical pyrolysis of humic substances from a Latosol (Typic Hapludox) under different land uses in Minas Gerais, Brazil, J. Anal. Appl. Pyrolysis 93 (2012) 120–128.
[23] C. Rumpel, A. Chabbi, N. Nunan, M.F. Dignac, Impact of landuse change on the molecular composition of soil organic matter, J. Anal. Appl. Pyrolysis 85 (2009) 431–434.
[24] J. Schellekens, G.G. Barbera, P. Buurman, G. Perez-Jorda, A. Martinez-Cortizas, Soil organic matter dynamics in Mediterranean A-horizons—the use of analytical pyrolysis to ascertain land-use history, J. Anal. Appl. Pyrolysis 104 (2013) 287–298.
[25] A.S. Grandy, G.P. Robertson, Land-use intensity effects on soil organic carbon accumulation rates and mechanisms, Ecosystems 10 (2007) 59–74.
[26] A.S. Grandy, M.S. Strickland, C.L. Lauber, M.A. Bradford, N. Fierer, The influence of microbial communities, management, and soil texture on soil organic matter chemistry, Geoderma 150 (2009) 278–286.
[27] A.S. Grandy, J.C. Neff, Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function, Sci. Total Environ. 404 (2008) 297–307.
[28] L. Berwick, P.F. Greenwood, R.J. Smernik, The use of MSSV pyrolysis to assist the molecular characterisation of aquatic natural organic matter, Water Res. 44 (2010) 3039–3054.
[29] A.S. Grandy, J.C. Neff, M.N. Weintrau, Carbon structure and enzyme activities in alpine and forest ecosystems, Soil Biol. Biochem. 39 (2007) 2701–2711.
[30] C.M. Kallenbach, S.D. Frey, A.S. Grandy, Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls, Nat. Commun. 7 (2016), 13630.
[31] F. Xu, Exergy and structural exergy as ecological indicators for the development state of the Lake Chaohu ecosystem, Ecol. Model. 99 (1997) 41–49.
[32] J. Huang, J. Zhan, H. Yan, F. Wu, X. Deng, Evaluation of the impacts of land use on water quality: a case study in the Chaohu Lake Basin, Sci. World J. 2013 (2013), 329187.
[33] T. Jiang, S. Huo, B. Xi, J. Su, H. Hou, H. Yu, X. Li, The influences of land-use changes on the absorbed nitrogen and phosphorus loadings in the drainage basin of Lake Chaohu, China, Environ. Earth Sci. 71 (2014) 4165–4176.
[34] J. Xiong, C. Lin, R. Ma, G. Zheng, The total P estimation with hyper-spectrum – a novel insight into different P fractions. Catena 187 (2020), 104309.
[35] C. Lin, R. Ma, J. Xiong, Can the watershed non-point phosphorus pollution be interpreted by critical soil properties? A new insight of different soil P states, Sci. Total Environ. 628–629 (2018) 870–881.
[36] M.S. Islam, M.K. Ahmed, M. Habibullah-Al-Mamun, S. Masunaga, Trace metals in soil and vegetables and associated health risk assessment, Environ. Monit. Assess. 186 (2014) 8727–8739.
[37] L. Chen, L. Liu, S. Qin, G. Yang, K. Fang, B. Zhu, Y. Kuzyakov, P. Chen, et al., Regulation of priming effect by soil organic matter stability over a broad geographic scale, Nat. Commun. 10 (2019) 5112.
[38] A. Kong, C. Gupta, M. Ferrari, M. Agostini, C. Bedin, A. Bouamrani, E. Tasciotti, R. Azencott, Biomarker signature discovery from mass spectrometry data, IEEE ACM Trans. Comput. Biol. Bioinf 11 (2014) 766–772.
[39] Ö. Gustafsson, F. Haghseta, C. Chan, J. MacFarlane, P.M. Gschwend, Quantification of the dilute sedimentary soot phase: implications for PAH speciation and bioavailability, Environ. Sci. Technol. 31 (1997) 203–209.
[40] M. Elmquist, Ö. Gustafsson, P. Andersson, Quantification of sedimentary black carbon using the chemothermal oxidation method: an evaluation of ex situ pretreatments and standard additions approaches, Limnol Oceanogr. Methods 2 (2004) 417–427.
[41] J. Xia, D.S. Wishart, Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis, Curr Protoc Bioinf 55 (2016), 14.10.11-14.10.91.
[42] R. Bivand, N. Lewin-Koh, E. Pebesma, E. Archer, A. Baddeley, N. Bearman, H.J. Bibiko, S. Brey, et al., Maptools: Tools for Handling Spatial Objects, R package version 0.9 5, 2019.
[43] E.J. Pebesma, Multivariable geostatistics in S: the gstat package, Comput. Geosci. 30 (2004) 683–691.
[44] R.C. Team, R: A Language and Environment for Statistical Computing, 2010.
[45] A.A. Malik, J. Puissant, K.M. Buckeridge, T. Goodall, N. Jehmlich, S. Chowdhury, H.S. Gweon, J.M. Peyton, et al., Land use driven change in soil pH affects microbial carbon cycling processes, Nat. Commun. 9 (2018) 3591.
[46] X. Xu, G. Pan, Z. Cao, Y. Wang, A study on the influence of soil organic carbon density and its spatial distribution in Anhui Province of China, Geogr. Res. 26 (2007) 1077–1086.
[47] C.Y. Jim, Urban soil characteristics and limitations for landscape planting in Hong Kong, Landsc. Urban Plann. 40 (1998) 235–249.
[48] X. Zang, P.G. Hatcher, A Py-GC-MS and NMR spectroscopy study of organic nitrogen in Mangrove Lake sediments, Org. Geochem. 33 (2002) 201–211.
[49] R. Conti, A.G. Rombolà, A. Modelli, C. Torri, D. Fabbri, Evaluation of the thermal and environmental stability of switchgrass biochars by Py-GC-MS, J. Anal. Appl. Pyrolysis 110 (2014) 239–247.
[50] J. Zhao, P.A. Peng, J. Song, S. Ma, G. Sheng, J. Fu, D. Yuan, Characterization of humic acid-like substances extracted from atmospheric falling dust using Py-GC-MS, Aerosol Air Qual. Res. 12 (2012) 83–92.
[51] M.I. Schnitzer, C.M. Monreal, G. Jandl, The conversion of chicken manure to bio-oil by fast pyrolysis. III. Analyses of chicken manure, bio-oils and char by Py-FIMS and Py-FDMS, J. Environ. Sci. Health, Part B 43 (2008) 81–95.
[52] L. Liu, X. Cheng, W. Zhao, Y. Wang, X. Dong, L. Chen, D. Zhang, W. Peng, Systematic characterization of volatile organic components and pyrolyzates from Camellia oleifera seed cake for developing high value-added products, Arab. J. Chem. 11 (2018) 802–814.
[53] M. Li, J.-H. Zhan, D. Lai, Y. Tian, X. Liu, G. Xu, Study on the evolution characteristic of intermediate during the pyrolysis of oil shale, J. Therm. Anal. Calorim. 130 (2017) 2227–2238.
[54] J. Zhao, P. Peng, J. Song, S. Ma, G. Sheng, J. Fu, Characterization of organic matter in total suspended particles by thermodesorption and pyrolysis-gas chromatography-mass spectrometry, J. Environ. Sci. 21 (2009) 62470–62475.
[55] Z. Miao, Y. Wan, Q. He, Z. Pei, X. Zhu, Pyrolysis Behaviors and Product Distribution of Shengli Lignite at Different Heating Rate and Final Temperature by TG-FTIR and Py-GC-MS, Energy Sources, Part As, 2019, pp. 3203–3215.
[56] W.A. Hartgers, J.S. S Damsté, J.W. de Leeuw, Geochemical significance of alkylbenzene distributions in flash pyrolysates of kerogens, coals, and asphaltenes, Geochem. Cosmochim. Acta 58 (1994) 1759–1775.
[57] J. Kaal, A. Martínez-Cortizas, K.G.J. Nierop, P. Buurman, A detailed pyrolysis-GC/MS analysis of a black carbon-rich acidic colluvial soil (Atlantic ranker) from NW Spain, Appl. Geochem. 23 (2008) 2395–2405.
[58] H. Zhou, C. Wu, A. Meng, Y. Zhang, P.T. Williams, Effect of interactions of biomass constituents on polycyclic aromatic hydrocarbons (PAH) formation during fast pyrolysis, J. Anal. Appl. Pyrolysis 110 (2014) 264–269.
[59] J.M. De la Rosa, S.R. Faria, M.E. Varela, H. Knicker, F.J. González-Vila, J.A. González-Pérez, J. Keizer, Characterization of wildfire effects on soil organic matter using analytical pyrolysis, Geoderma 191 (2012) 24–30.
[60] A.P. Deshmukh, B. Chefetz, P.G. Hatcher, Characterization of organic matter in pristine and contaminated coastal marine sediments using solid-state 13C NMR, pyrolytic and thermochemolytic methods: a case study in the San Diego harbor area, Chemosphere 45 (2001) 1007–1022.
[61] O. Pisani, M.L. Haddix, R.T. Conant, E.A. Paul, M.J. Simpson, Molecular composition of soil organic matter with land-use change along a bi-continental mean annual temperature gradient, Sci. Total Environ. 573 (2016) 470–480.
[62] O.E. Sala, F. Stuart Chapin, J.J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L.F. Huenneke, et al., Global biodiversity scenarios for the year 2100, Science 287 (2000) 1770–1774.
[63] M.W.I. Schmidt, Carbon budget in the black, Nature 427 (2004) 305–307.
[64] C.A. Masiello, E.R.M. Druffel, Black carbon in deep-sea sediments, Science 280 (1998) 1911–1913.
[65] J. Lehmann, C.M. Hansel, C. Kaiser, M. Kleber, K. Maher, S. Manzoni, N. Nunan, M. Reichstein, et al., Persistence of soil organic carbon caused by functional complexity, Nat. Geosci. 13 (2020) 529–534.
[66] J.Y. Yang, F. Yu, Y.C. Yu, J.Y. Zhang, R.H. Wang, M. Srinivasulu, V.I. Vasenev, Characterization, source apportionment, and risk assessment of polycyclic aromatic hydrocarbons in urban soil of Nanjing, China, J. Soils Sediments 17 (2017) 1116–1125.
[67] S. Liu, X. Xia, Y. Zhai, R. Wang, T. Liu, S. Zhang, Black carbon (BC) in urban and surrounding rural soils of Beijing, China: spatial distribution and relationship with polycyclic aromatic hydrocarbons (PAHs), Chemosphere 82 (2011) 223–228.
[68] J.L. Edmondson, I. Stott, J. Potter, E. Lopez-Capel, D.A. Manning, K.J. Gaston, J.R. Leake, Black carbon contribution to organic carbon stocks in urban soil, Environ. Sci. Technol. 49 (2015) 8339–8346.
[69] J.B. Paroissien, T.G. Orton, N.P.A. Saby, M.P. Martin, C.C. Jolivet, C. Ratie, G. Caria, D. Arrouays, Mapping black carbon content in topsoils of central France, Soil Use Manag. 28 (2012) 488–496.
[70] H. Fu, H. Liu, J. Mao, W. Chu, Q. Li, P.J.J. Alvarez, X. Qu, D. Zhu, Photochemistry of dissolved black carbon released from biochar: reactive oxygen species generation and phototransformation, Environ. Sci. Technol. 50 (2016) 1218–1226.
[71] X. Qu, H. Fu, J. Mao, Y. Ran, D. Zhang, D. Zhu, Chemical and structural properties of dissolved black carbon released from biochars, Carbon 96 (2016) 759–767.