

Spatial patterns and environmental functions of dissolved organic matter in grassland soils of China

Received: 12 December 2023

Accepted: 19 July 2024

Published online: 28 July 2024



Peng Zhou^{1,2}, Long Tian^{1,2}, Nigel Graham³, Shian Song¹, Renzun Zhao⁴, Muhammad Saboor Siddique^{1,2}, Ying Hu^{2,5}, Xianyong Cao⁶, Yonglong Lu⁷ , Menachem Elimelech⁸  & Wenzheng Yu¹ 

Soil dissolved organic matter (DOM) is crucial to atmospheric, terrestrial and aquatic environments as well as human life. Here, by characterizing DOM from 89 grassland soils throughout China, we reveal the spatial association between DOM geochemistry in the dry season vs annual ecosystem exchange and cancer cases. The humic-like and high molecular weight (3.4–25 kDa) fractions with lower biodegradability, decline from the northern to the southern regions of China, and are correlated with lower soil respiration and net ecosystem productivity at the continental scale. The <1.2 kDa and proteinaceous fractions could serve as a geographical indicator of nasopharyngeal cancer incidence and mortality, while the 3.4–25 kDa and humified fractions are potentially associated with pancreatic cancer cases ($P < 0.05$). Our findings highlight that exploiting the environmental functions of soil DOM and mitigating the negative impacts are necessary, and require actions tailored to local soil DOM conditions.

Soil-dissolved organic matter (DOM) represents an important portion of terrestrial organic matter, playing a role in mediating energy and nutrient transfer between land, water, and the atmosphere^{1,2}. In terrestrial ecosystems, DOM is subject to mobilization, deposition, and microbial degradation and can be converted into CO₂, microbial products, and partially degraded molecules^{3–5}. The transformation and mobilization of terrestrial DOM play different roles in regulating nutrient availability and microbial activity, balancing ecosystem carbon, and altering the ambient environment⁶. Intrinsically, these ecosystem functions would vary greatly with ecosystem types and influence wider processes, because of the diverse and complex molecular properties of DOM.

The input of terrestrial DOM into fluvial systems alters water quality and aquatic ecosystems markedly, ultimately impacting drinking water production and human health. The intruded DOM directly increases the organic carbon content and changes the DOM quality in water bodies^{7,8}. Moreover, such DOM in catchments supplying source water to drinking water treatment plants poses further challenges. First, it affects the treatment efficiency of coagulation, adsorption, oxidation, and membrane filtration processes that are used in water treatment. Second, it would lead to health concerns. The prevalence of some endemic diseases of unknown etiology (e.g., kidney disease) was related to the special characteristics of natural

¹State Key Laboratory of Environmental Aquatic Chemistry, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ³Department of Civil and Environmental Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. ⁴Department of Civil, Architectural and Environmental Engineering, North Carolina Agricultural and Technical State University, Greensboro, North Carolina, USA. ⁵State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China. ⁶Group of Alpine Paleoecology and Human Adaptation (ALPHA), State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources (TPESER), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China. ⁷State Key Laboratory of Marine Environmental Science and Key Laboratory of the Ministry of Education for Coastal Wetland Ecosystems, College of the Environment and Ecology, Xiamen University, Fujian 361102, China. ⁸Department of Chemical and Environmental Engineering, Yale University, New Haven, CT 06520-8286, USA. ✉ e-mail: yllu@xmu.edu.cn; yllu@rcees.ac.cn; menachem.elimelech@yale.edu; wzyu@rcees.ac.cn

organic matter in local environments^{9,10}. Because these compounds may be either potentially toxic or interact with harmful chemicals and pollutants^{11,12}, which can enter and remain in human tissues¹³. Additionally, during drinking water disinfection (e.g., chlorination), byproducts (e.g., trihalomethanes and haloacetic acids) are formed by the unavoidable reactions with DOM, which may have toxicity and carcinogenicity^{14,15}. Since humans can suffer from long-term exposure to DOM in drinking water¹⁶, there may be a link between source DOM and human health, yet this is ill-defined and remains poorly investigated.

Terrestrial DOM is, in turn, susceptible to environmental conditions. The input of plant materials determines the abundance and quality of organic matter and also alters microbial environments in soils^{17,18}. Edaphic properties, such as pH¹⁹, water content²⁰, metal/metal oxides concentration²¹, and clay fraction²², can alter the soil microenvironment and molecular mobility and bioavailability. Moreover, climate (e.g., temperature and rainfall) affects the transport and chemical reactions of molecules and regulates the interactions among soil, plants, and microbes^{4,18}. In view of this, the fate of DOM and the subsequent environmental impacts, vary as a function of the combined effects of ecosystem properties. To date, however, the underlying control-response mechanisms related to the ecosystems and soil DOM, as well as their environmental impacts, are poorly understood, especially under the ongoing climate change scenario.

Grasslands comprise ~40% of the world's ice-free land surface and store over one-third of terrestrial carbon globally²³. They represent vast soil organic carbon stocks with dissolved organic carbon (DOC) concentrations comparable to the forests and croplands²⁴. In China, grassland landscape also accounts for approximately 40% of the total land area, which varies in climate, plant biome, soil type and topography. Consequently, the DOM in widespread grassland areas should have profound impacts on terrestrial and aquatic environments. Moreover, due to the connections between soil and drinking water sources, the DOM may be potentially associated with human activities and health^{10,25,26}. A few studies have previously considered the quantity or quality of grassland soil DOM in China, but mainly at a regional scale^{27,28}. Systematic studies of the DOM geochemistry and the controlling factors at large spatial scales are scarce. Crucially, there are gaps in our understanding of ecosystem roles and environmental impacts of the ubiquitous terrestrial DOM, especially its potential health effects.

In this study, we ask two questions: (i) What are the compositional features and potential environmental functions of soil DOM in large-scale grassland ecosystems? (ii) How do environmental factors, particularly climate variables, drive the spatial patterns of DOM at the continental scale? To address these questions, we conducted a nationwide survey which characterized DOM in 89 grassland soil samples across 30 provinces of China in the dry season (Supplementary Figs. S1, S2). DOM is analyzed for fluorescence components, molecular weight (MW) distribution, and molecular composition. Given the concerns regarding drinking water safety, the disinfection byproduct formation potential (DBP-FP) and chlorine reactivity of DOM are examined. We then assess the spatial patterns of DOM quantity and quality, and their relationships to ecosystem carbon exchange, under the control of environmental factors. Furthermore, the DOM is associated with local cancer incidence and mortality rates to explore the potential connection between terrestrial-derived DOM and human health.

Results and discussion

Considerable storage of DOM in grassland soils

Four fluorescent components (C1–C4) were identified from the soil DOM in grasslands (Fig. 1a, Supplementary Fig. S3, and Table S1). Details on the fluorescent components are described in

Supplementary Text S1. The humic-like (C1, C2, and C3) and medium-to-high molecular weight (1.2–25 kDa) components comprised the major grassland soil DOM (Fig. 1b, c). The mean concentrations of dissolved organic carbon (DOC) and nitrogen (DON), trihalomethane precursors (THM-FP) and haloacetic acid precursors (HAA-FP) in grassland soils were 199.9 [15.3–804.5], 14.3 [0.4–68.4], 33.7 [2.6–123.8] and 4.3 [0.3–16.9] mg·kg⁻¹, respectively (Fig. 1d and Supplementary Fig. S4). The DOC and DON concentrations are in good agreement with the previous regional records in China's grasslands^{27,29}. Moreover, the DOM quantity was largely related to the abundance of high molecular weight (HMW; 3.4–25 kDa) and humic-like (C1–C3) compounds ($P < 0.05$) (Fig. 1f), and nearly 40% of the DOC and DON variations can be explained by the fluorescent and size fractions (Fig. 1g).

The grassland soils display higher DOC abundance than many forest and cropland soils²⁴, and have a similar molecular diversity of DOM as compared to the other ecosystems (Supplementary Fig. S5)^{30,31}. The fluorescent compounds in grassland soils were identified in diverse soil and aquatic ecosystems^{32–34}. Furthermore, lignin-, tannin-, and condensed aromatic-like compounds, mainly with intermediate O/C and H/C were universal in different grasslands (Supplementary Fig. S6). These compounds are prevalent in various ecosystems and are relatively persistent in the environments^{30,31,35,36}. Consequently, grassland soils are enormous DOM reservoirs, which have high potential influences on wider ecosystems and processes. However, the impacts of soil DOM may have large spatial differences because of the wide range of DOM characteristics and chemistry of water extracts across grasslands (Fig. 1b–e and Supplementary Fig. S7).

Compositional signature of grassland soil DOM

The humic-like (C1 + C2 + C3) (accounted for over 70% of the fluorescent components on average) and high molecular weight (HMW, 3.4–25 kDa) compounds from plant sources dominate the DOM composition in grassland soils, compared with the microbial-derived proteinaceous compounds (C4) and low (LMW, <1.2 kDa) and very high molecular weight (VHMW, >25 kDa) compounds (Fig. 1b, c and Supplementary Fig. S8). Accordingly, CHO and CHNO compounds, such as lignin-, tannin-, and condensed aromatic-like compounds, also comprised the majority of the molecular composition (Supplementary Fig. S5). Compared with the DOM in some other soil ecosystems, such as forest, permafrost and arid/semi-arid cropland³⁷, DOM in grassland soils displayed high average HIX and FI (>1.5) values concurrently (Fig. 1e), which was consistent with the observations of surface grassland soil DOM at regional scale^{38,39}. This might be because both microbial activity and plant input are high in grassland soils, while compounds from the microbial source contribute more to the organic matter composition⁴⁰.

The composition of grassland soil DOM was between two different patterns. The one is featured by higher humic-like and HMW (3.4–25 kDa) fractions, and the other is characterized by the greater protein-like and LMW (<1.2 kDa) fractions (Fig. 2a–d and Supplementary Fig. S9). The relative abundance (%) of humic-like components (C1, C2, and C3) was positively correlated with the percentage of HMW compounds and HIX ($P < 0.05$) (Supplementary Fig. S10). Moreover, the humic-like and HMW fractions were mainly positively associated with lignin-, tannin- and condensed aromatic-like compounds with H/C < 1 and O/C > 0.5 (Fig. 2e, f and Supplementary Fig. S11), which was largely produced from the breakdown of plant litter in surface soils^{41,42}. In contrast, the protein-like (C4) fraction was positively correlated with the relative abundance of VHMW and LMW compounds and BIX ($P < 0.05$) (Supplementary Fig. S10). The protein-like and LMW fractions were related to compounds with intermediate O/C and medium-to-high H/C, especially the small amino sugar- and lipid-like compounds, which were less aromatic and derived from biological processes (Supplementary Fig. S11)³⁵.

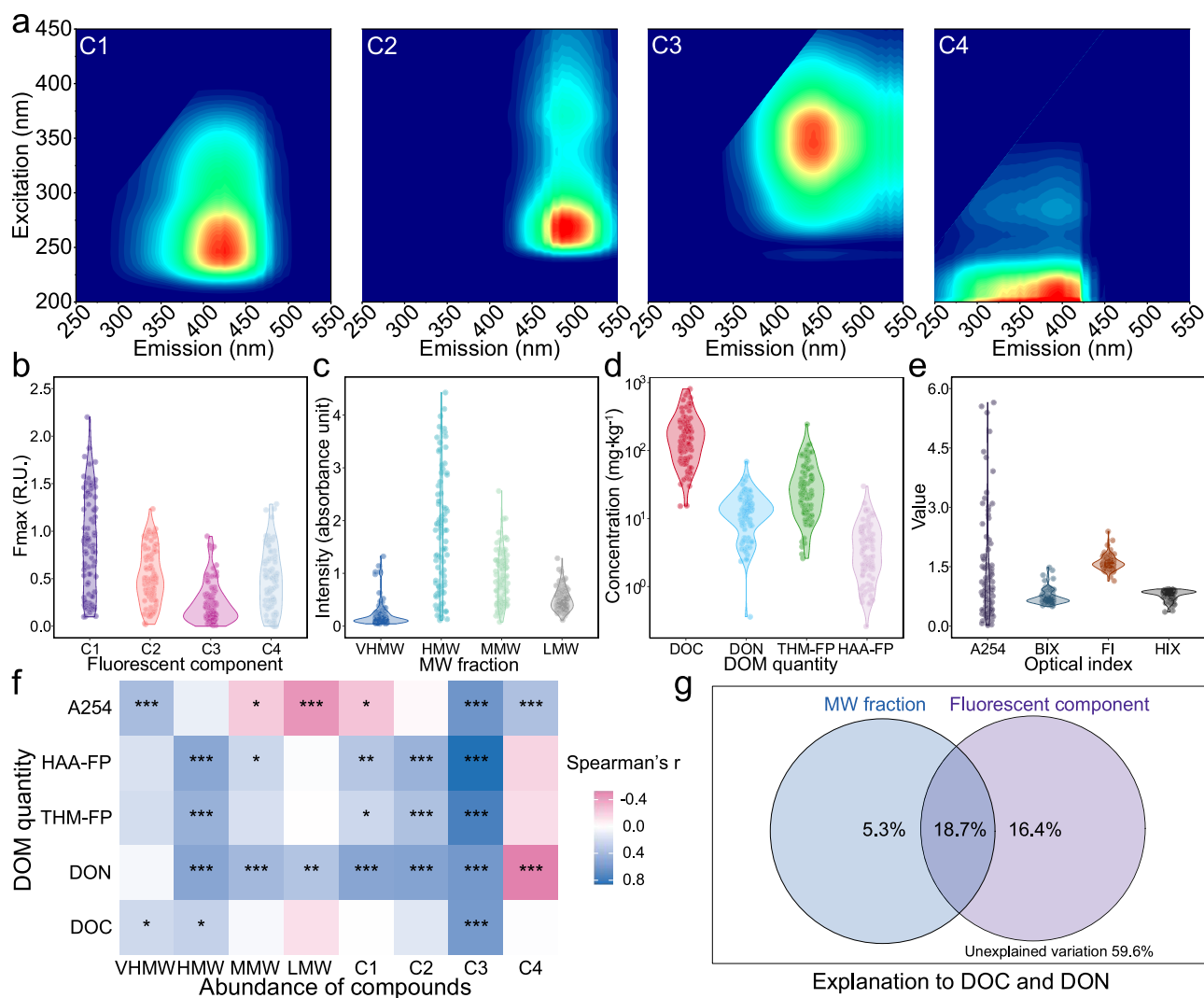


Fig. 1 | The storage of different dissolved organic matter (DOM) species in grassland soils. a Spectral properties of four fluorophores (C1, C2, C3, and C4) identified by parallel factor analysis. **b** Intensity of the four fluorescent components in Raman unit (R.U.). **c** Cumulative intensity (absorbance unit) of the four fractions with different molecular weights. VHMW, 25–100 kDa; HMW, 3.4–25 kDa; MMW, 1.2–3.4 kDa; LMW, <1.2 kDa. **d** Concentrations (mg·kg⁻¹·soil) of dissolved organic carbon (DOC), dissolved organic nitrogen (DON), trihalomethane precursors (THM-FP), and haloacetic acid precursors (HAA-FP). **e** UV absorbance at 254 nm

(A254), biological index (BIX), fluorescence index (FI), and humification index (HIX) of DOM. **f** Spearman correlation between DOC, DON, disinfection byproduct precursors, and A254 vs different DOM compounds. **g** Variation partitioning analysis for DOC and DON. Numbers in each circle are the unique effects of size fraction and fluorescent component, respectively, and the number in the overlapping area of the two circles is the shared effects of the two factors. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

Spatial features of DOM geochemistry

The northern and southern regions of China varied in DOM geochemistry. The northern regions of China displayed higher 3.4–25 kDa and humic-like (C1 + C2 + C3) fractions, greater humification degree (HIX), and higher concentrations of DOC, DON, and DBP precursors (DBP-FP) (Fig. 3a, c, and Supplementary Figs. S12–S14). These observations indicate a greater DOM abundance in these regions, particularly higher proportions of allochthonous humified compounds with high aromaticity and MW, including lignin-, tannin-, and condensed aromatic-like molecules with $H/C < 1$ and $O/C > 0.5$ (Supplementary Fig. S15). By contrast, higher levels of LMW, MMW, C4, biological index (BIX) and fluorescence index (FI) were apparent in southern China (Fig. 3b, d and Supplementary Figs. S12, S13), which had less DOM but was more enriched in autochthonous <3.4 kDa and proteinaceous compounds, especially the less oxidized protein/amino sugar- and lipid-like compounds with $H/C > 1.5$ (Supplementary Fig. S15).

The distinct patterns of the fluorescent and molecular weight composition of soil DOM for northern and southern China imply

divergent effects on local or downstream ecosystems and processes. For one thing, the DOM in different regions may play varied roles in ecosystem exchange. For another, it suggests distinct influences on aquatic environments after being released into waters during rainfall and erosion events⁷. The DOM compounds in northern China are more recalcitrant^{14,43}, and may impose longer negative effects on local and downstream water chemistry and drinking water production due to the longer persistence in the environment. In contrast, the more labile DOM fractions in southern China are more readily mineralized and would increase carbon release. They have less persistent impacts on aquatic systems and water sanitation but can increase labile nutrients in waters and improve CO₂ outgassing from the waters⁴⁴. In addition, DOM in the northwestern and Qinghai-Tibetan regions, was more reactive in terms of trihalomethanes (THMs) and haloacetic acids (HAAs) formation, as displayed by the notably higher chlorine reactivity (SDBP) (Fig. 3e, f). Hence, these regions, with water towers for Asia⁴⁵, can be sources of highly reactive DBP precursors to downstream waters.

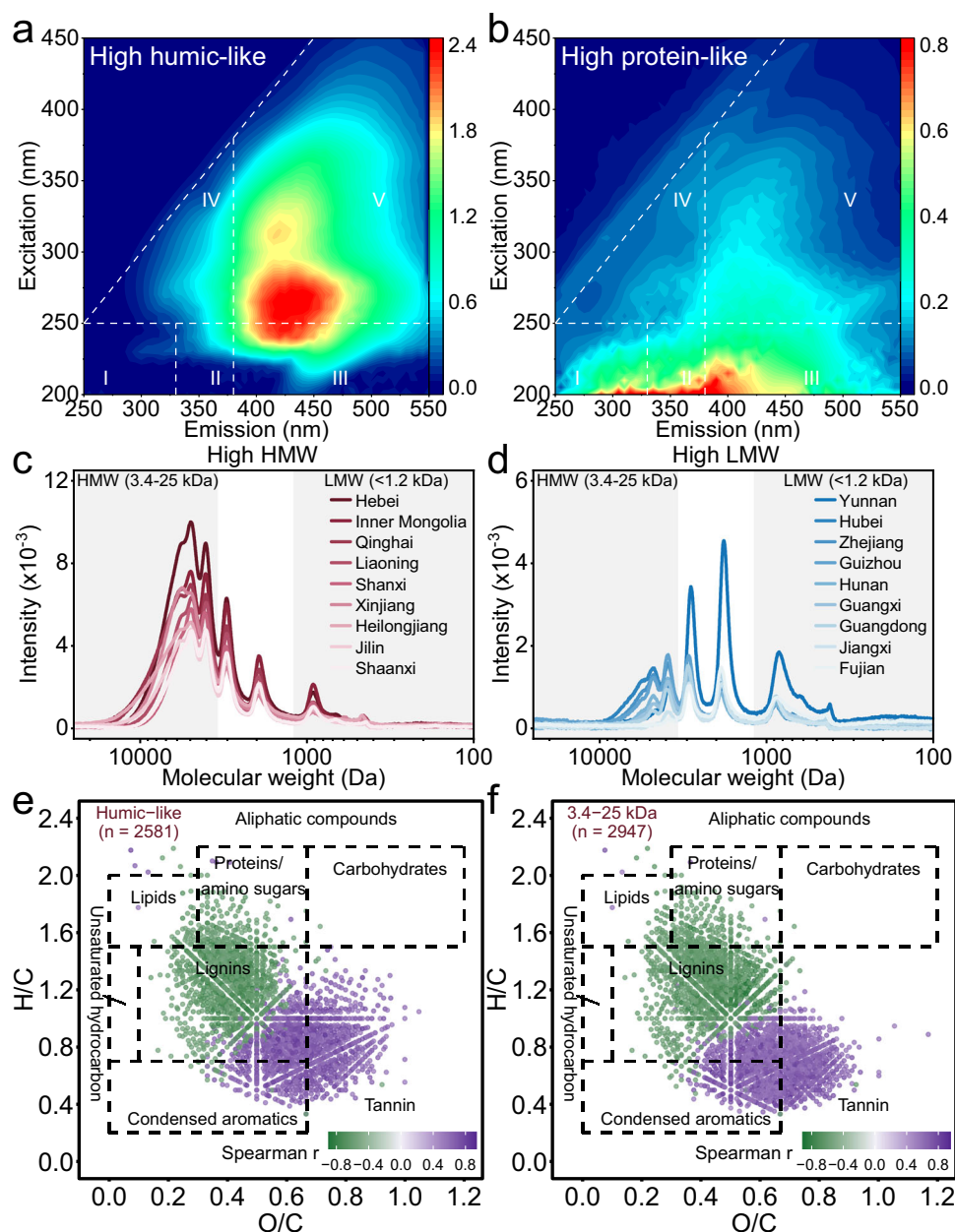


Fig. 2 | Two distinct signatures of the composition of dissolved organic matter (DOM) in grassland soils. a, b Fluorescence spectra of the DOM with **a** high humic-like and **b** high protein-like fractions. Region I, tyrosine protein-like; Region II, tryptophan protein-like; Region III, fulvic acid-like; Region IV, soluble microbial byproduct-like; and Region V, humic acid-like. **c, d** Apparent molecular weight distribution of the DOM with **c** high HMW and **d** high LMW fractions. **e, f** Significant

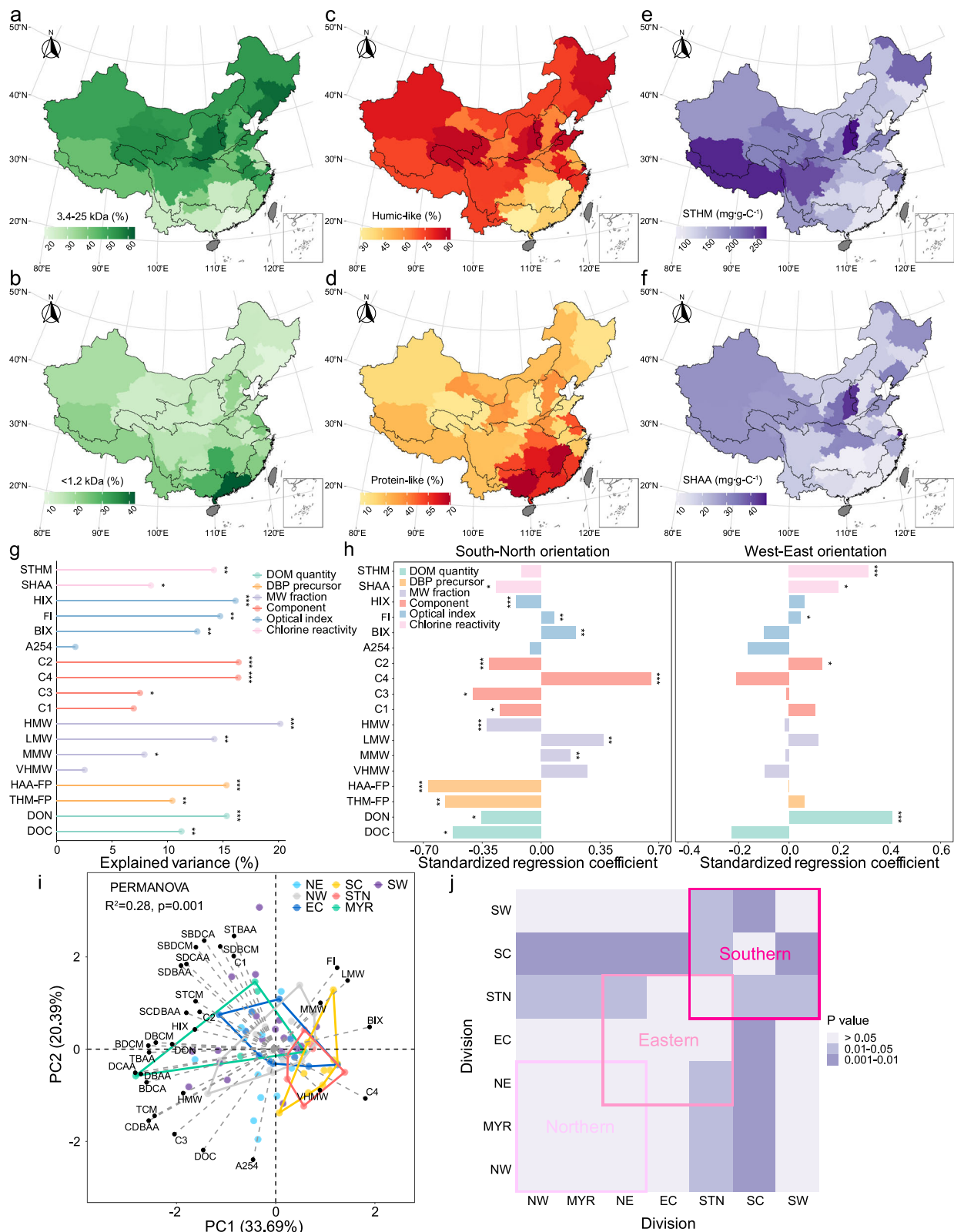
Spearman's rank correlations ($P < 0.05$) between the relative abundance of individual molecules vs **e** humic-like (C1 + C2 + C3) and **f** 3.4–25 kDa fractions. The green and purple colors of the circles indicate negative and positive correlation coefficients, respectively. The dashed lines in van Krevelen diagrams separate the molecules into different compound categories. HMW, 3.4–25 kDa; LMW, <1.2 kDa.

Multiple linear regression further demonstrated significant variations of DOM geochemistry with its geographical location (Fig. 3g and Supplementary Fig. S16). Fluorophores, size fractions, HIX, BIX, and DOM quantity, displayed a divide in the South-North orientation ($P < 0.05$), while the dissolved nitrogen and SDBP varied in the West-East orientation ($P < 0.05$) (Fig. 3h and Supplementary Fig. S16). The Yangtze River basin serves as a watershed where the characteristics of DOM varied remarkably in both South-North and West-East orientation. Interestingly, DOM in provinces along the Middle Route of the South-North Water Transfer Project (Supplementary Fig. S17) varied gradually and displayed greater <1.2 kDa and protein-like fractions with lower aromaticity (HIX) than the surrounding provinces (Fig. 1b, d and Supplementary Fig. S14).

The DOM quantity, DBP precursors, HMW, and humic-like fractions, showed negative PC1 loadings in the principal component analyses, in contrast with the C4, LMW, MMW, VHMW, FI, and BIX (Fig. 3i). The SC and STN regions exhibited positive PC1 scores, indicating more microbial features of DOM in these regions. In contrast, the MYR and NE regions were mainly featured by the higher abundance of plant-derived DOM. Indeed, grassland soil DOM in the southern regions (STN and SC) differed significantly from that in the northern regions (NW, MYR, and NE) (Fig. 3j).

Roles in ecosystem carbon exchange

After incubation for 31 days, soil DOM with higher humic-like and HMW fractions (mainly from the northern regions) displayed 21–35%



lower biodegradability ($P < 0.01$), compared to that with higher protein-like and LMW fractions (mainly from the southern regions) (Fig. 4a). Furthermore, the higher humic-like and HMW fractions led to over 20% higher stable pools for the DOM ($P < 0.05$) (Supplementary Fig. S18). In contrast to the protein-like fractions and compounds with $O/C < 0.5$ and $H/C > 1$, the humic-like and HMW fractions, and compounds with higher O/C and lower H/C , were negatively correlated

with the biodegradation of DOM but positively correlated with the stable fractions of DOM ($P < 0.05$) (Supplementary Figs. S19, S20). As a result, the biological decomposition of soil DOM was significantly faster because of the greater protein-like and LMW composition ($P < 0.05$) (Fig. 4b).

On this basis, the association between the grassland soil DOM observed in the dry season and mean annual ecosystem carbon were

Fig. 3 | Spatial features of soil-dissolved organic matter (DOM) in China's grasslands. **a–f** Spatial distribution of the relative abundance of **a** 3.4–25 kDa, **b** <1.2 kDa, **c** humic-like (C1 + C2 + C3), and **d** protein-like (C4) compounds, and the chlorine reactivity of **e** trihalomethane (STHM) and **f** haloacetic acid (SHAA) precursors. Values in maps are the average of sites in each province. The black lines indicate the boundaries of the nine river basins of China. The gray area denotes no data observed. **g, h** Explanation of DOM geochemistry by geographical location (South-North and West-East orientation) shown by **g** regression model variance and **h** standardized regression coefficients. **i** Spatial variations in DOM geochemistry shown by principal component analysis with permutational multivariate analysis of variance (PERMANOVA). The seven regions were displayed by seven colors,

including the east coast (EC), middle of Yellow River (MYR), northeast (NE), northwest (NW), south coast (SC), South-North Water Transfer (STN), southwest (SW) regions. The polygons indicate the groups of different regions. **j** The significance (*P* value) of geochemical differences among the seven regions in PERMANOVA. THM-FP and HAA-FP are trihalomethane (including TCM, BDCM, and DBCM) and haloacetic acid (including DCAA, BDCA, CBDAA, DBAA, and TBAA) precursors, respectively, and details are listed in Supplementary Table S3. DOC, dissolved organic carbon; DON, dissolved organic nitrogen; VHMW, 25–100 kDa; HMW, 3.4–25 kDa; MMW, 1.2–3.4 kDa; LMW, <1.2 kDa; A254, UV absorbance at 254 nm; BIX biological index, FI fluorescence index, HIX humification index. **P* < 0.05; ***P* < 0.01; ****P* < 0.001.

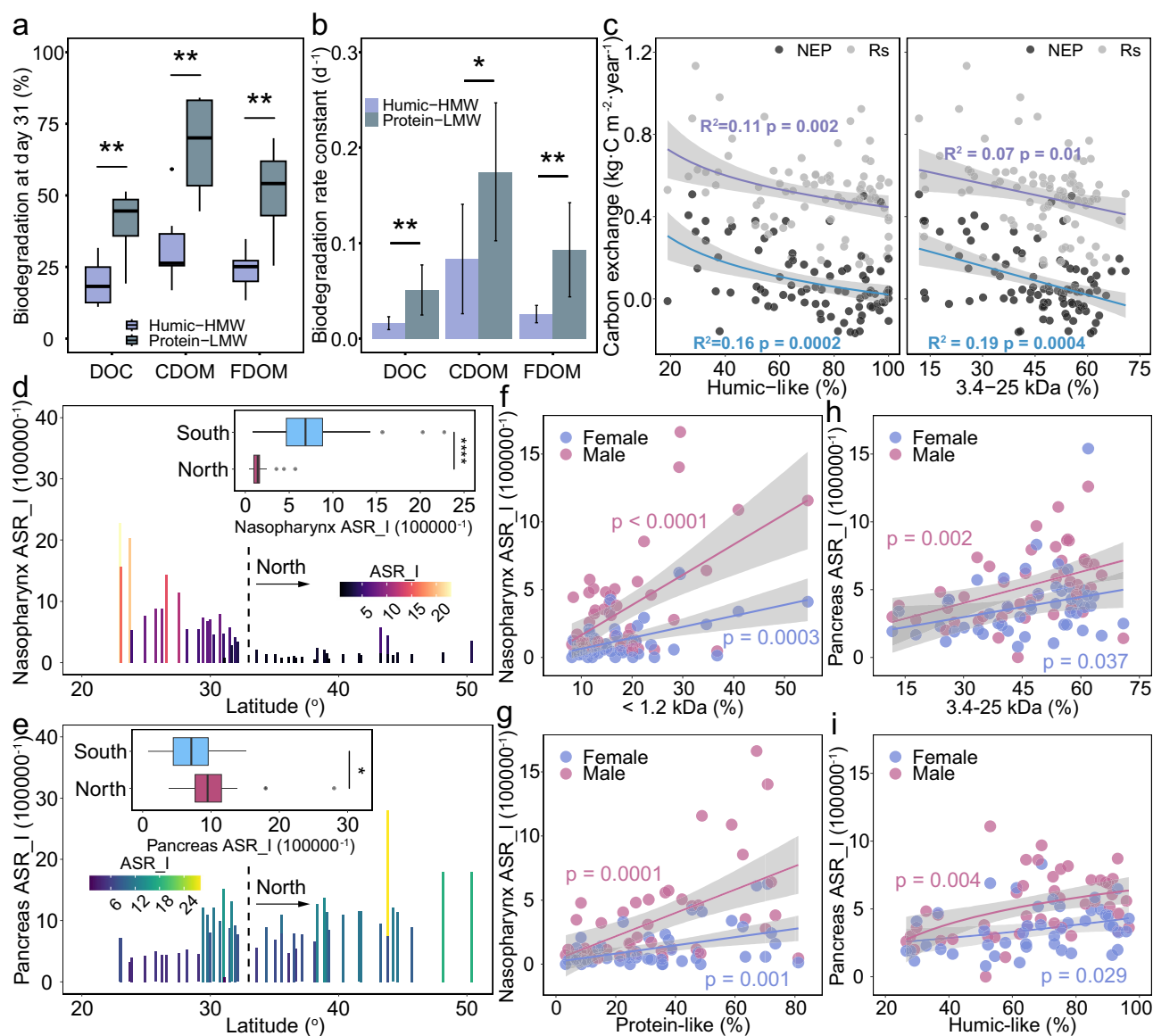


Fig. 4 | Association between dissolved organic matter (DOM) composition vs ecosystem exchange and human health. **a** Biodegradation level (%) at day 31 and **b** average biodegradation rate (d^{-1}) of DOM during incubation. DOC dissolved organic carbon, CDOM chromophoric DOM, FDOM fluorophoric DOM. The value of the biodegradation rate constant is shown by mean $\pm$ standard error in the bar plot. **c** Comparisons between Rs and NEP vs the percent relative abundance (%) of humic-like and 3.4–25 kDa components. Rs soil respiration, NEP net ecosystem productivity. **d, e** Spatial distribution and north-south difference of **d** nasopharynx and **e** pancreas cancer incidence ($100,000^{-1}$). The north represents the sites with

latitudes above 33° , and the south represents those with latitudes lower than 33° . The Wilcoxon test was applied to examine the difference between the north and south. **f, g** Relationships between nasopharynx cancer incidence and **f** <1.2 kDa and **g** protein-like fractions. **h, i** Relationships between pancreatic cancer incidence and **h** 3.4–25 kDa and **i** humic-like fractions. ASR_I, age-standardized rate of cancer incidence. The fitting line is based on regression analysis (including linear, logarithmic, and exponential regression), with a line indicating mean model fit $\pm 95\%$ confidence interval. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

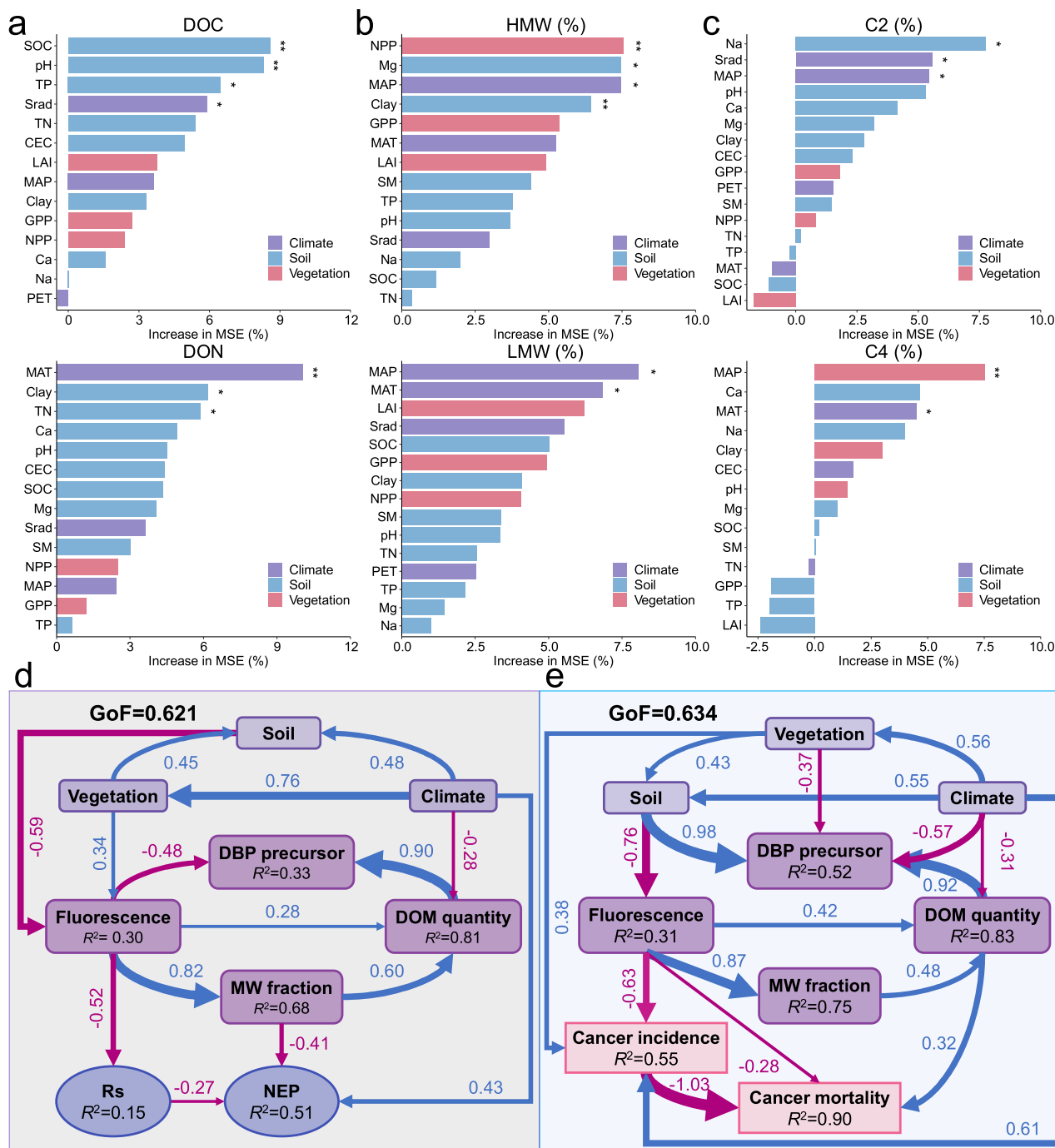


Fig. 5 | Environmental drivers for the geochemistry and associated impacts of soil-dissolved organic matter (DOM). **a–c** The importance of environmental predictors for **a** DOM quantity, **b** size fraction, and **c** fluorescence composition based on random forest regression analysis. An increase in the mean square error (MSE) denotes the higher importance of environmental variables. All models were significant at $P < 0.01$. DOC dissolved organic carbon, DON dissolved organic nitrogen, HMW 3.4–25 kDa, LMW <1.2 kDa, MAT mean annual temperature, MAP mean annual precipitation, Srad surface solar radiation, PET potential evapotranspiration, LAI leaf area index, NPP net primary production, GPP gross primary production, SM soil moisture, Clay soil clay fraction, pH soil pH, CEC cation exchange capacity of soil, Na, Mg, and Ca soil exchangeable sodium, potassium,

and calcium, respectively, SOC soil organic carbon, TN soil total nitrogen, TP soil total phosphorus. **d, e** Partial least-squares path model (PLS-PM) showing the relationships between DOM vs **d** ecosystem carbon exchange and **e** local cancer incidence and mortality, under the influence of environmental variables. Rs soil respiration, NEP net ecosystem productivity. A detailed description of the latent variables in the model is available in Supplementary Table S2. The purple and red lines in PLS-PM denote significant positive and negative relationships between variables, respectively. PLS-PM are built after 999 bootstraps and only significant ($P < 0.05$) paths are shown. Numbers in the arrows are standard path coefficients, and the determination coefficients (R^2) are the variance explained by the model. GoF goodness of fit. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

examined to reveal the roles of the DOM in ecosystem exchange at the continental scale. DOM quantity is positively associated with soil organic matter ($P < 0.01$) (Supplementary Fig. S21). In this respect, although DOM is a small portion of organic matter in soil²⁴, it is still

associated with soil carbon storage. The 3.4–25 kDa and humic-like fractions were negatively associated with soil respiration (Rs) and net ecosystem productivity (NEP) ($P < 0.05$) (Fig. 4c and Supplementary Fig. S22). In contrast, the relative abundance of <1.2 kDa and protein-

like compounds were positively correlated with these ecosystem carbon ($P < 0.01$) (Supplementary Fig. S22). Consequently, given the spatial differences in the molecular composition (Fig. 3), the grassland soil DOM present in the dry season plays distinct roles in annual ecosystem carbon exchange in the northern and southern regions of China. The LMW and proteinaceous portions are associated with less aromatic and persistent molecules that are more readily available to microbes (Supplementary Figs. S11, S20). Consequently, preferential mineralization of these labile components during microbial metabolism can accelerate carbon emission (Fig. 4b), and balance the higher ecosystem productivity in the southern and eastern regions of China⁴⁶. However, the HMW and humic-like components originating from plant sources are the products of atmospheric carbon with a more recalcitrant nature to biodegradation (Supplementary Figs. S11, S20). These compounds can suppress carbon loss (Fig. 4a), and thus conserve ecosystem productivity in northern and western China⁴⁶.

Potential linkage with human health

The age-standardized rate of cancer incidence (ASR_I) and mortality (ASR_M) of the nasopharynx, pancreas, and kidney and urinary organs cancers varied with the latitude, and differed significantly from the southern to the northern regions of China ($P < 0.05$) (Fig. 4d, e and Supplementary Figs. S23, S24), in line with the spatial distribution of grassland soil DOM composition in the dry season (Fig. 3). Regression analysis further demonstrated that higher LMW and protein-like fractions were correlated with higher nasopharynx cancer incidence and mortality ($P < 0.01$), while HMW and humic-like components were positively related to the pancreas as well as kidney and urinary organs cancer cases ($P < 0.05$) (Fig. 4h, i and Supplementary Figs. S23, S24).

The associations could be partly ascribed to two reasons: (i) Some organic compounds, such as hydrophilic proteinaceous and large-sized humic compounds, might be harmful to human cells, as reported in many bioassays^{11,25,26,47}; (ii) DOM interacted with other toxic chemicals or pollutants, such as arsenic and fluoride, may cause adverse health outcomes^{12,48,49}. Previous studies also related the unique DOM compounds (e.g., humic-like) in natural environments to endemic diseases (e.g., chronic kidney disease)^{9,10}. Moreover, some particular organic components (e.g., tryptophan and tyrosine protein-like) were found present in pathological tissues^{13,50–53}. Therefore, since the causative factors of endemic nasopharynx, pancreas, and kidney and urinary organ cancers are largely unknown, the significant correlations suggested that the soil DOM could play an indicative role. Moreover, the endemic cancers of unknown etiology were related to the DOM in source waters^{9,12}. Similarly, we found a significant correlation between grassland soil DOM and the organic carbon concentration of local tap water ($P < 0.05$) (Supplementary Fig. S25). Thus, exposure to organic carbon (0–7 mg-C-L⁻¹) in drinking water may be partially responsible for the potential associations between soil DOM and cancer cases in China¹⁶.

Given the connections between soil, source water, and drinking water, the associations between soil DBP precursors and cancer cases were examined. Trihalomethane precursors (THM-FP) were positively correlated with liver cancer cases, while haloacetic acid precursors (HAA-FP) were positively correlated with additional kinds of cancers concerning pancreas, kidney and urinary organs and bladder ($P < 0.05$) (Supplementary Figs. S26, S27). Exposure to toxic DBPs (including THMs and HAAs) in drinking water was indeed associated with increased risks of liver, pancreatic, kidney, and bladder cancers previously^{16,54–56}. Thus, the formation of harmful byproducts in drinking water may be another potential pathway linking the DOM to local human health.

Lastly, we must emphasize that our findings, based on statistical significance, only highlight the potential association between soil DOM and human health rather than proposing that DOM is the pathogenic factors. In this context, the grassland DOM geochemistry at the

continental scale in the dry season could serve as a geographical indicator of endemic cancers. Moreover, the results provide new insights into the screening and indication of environmental risk factors for endemic diseases of unknown etiology, and further studies are required to reveal the underlying mechanisms.

Environmental controls over the DOM

Random forest (RF) analysis was first conducted to reveal the importance of environmental factors to predict DOM features. Edaphic properties, including soil organic carbon (SOC) and total nitrogen (TN), were key predictors for DOC, DON, THM-FP, and HAA-FP ($P < 0.05$) (Fig. 5a, and Supplementary Fig. S28), while DOM composition (LMW, HMW, C2, and C4) was primarily explained by climate parameters (Fig. 5b, c). The climate, vegetation, and soil factors largely drive the variations in soil DOM across China (Fig. 5d, e). Among them, both climate (temperature, MAT; precipitation, MAP; and solar radiation, Srad) and soil (soil moisture, pH, clay fraction, and Ca content) parameters had negative effects on DOM quantity (DOM) and quality (fluorescence composition, FC; and molecular weight fractions, MW) (Fig. 5d, e and Supplementary Fig. S29). Compared to the southern China region, the northern China region overall had relatively lower MAT, MAP, Srad, net primary production, and soil moisture and clay fraction but higher soil mineral ions^{57–60}, which contribute to the accumulation of HMW and humic-like fractions (Supplementary Fig. S29). Consequently, the differences in these environmental factors were responsible for the greater humic-like and HMW composition for the northern regions and the greater protein-like and LMW fractions for the southern regions. Also, the distinct DOM properties along the South-North Water Transfer Project may partly result from the altered climate (evaporation and precipitation) at the regional scale⁶¹.

The DOM is regulated by environmental factors, and its variation, in turn, controls the roles in ecosystem exchange and is linked with local human health. On the one hand, the grassland soil DOM in the dry season and environmental variables explained 15 and 51% of the variance (R^2) of annual soil respiration (Rs) and net ecosystem productivity (NEP), respectively (Fig. 5d). Under the control of environmental drivers, Fluorescence and MW fraction mainly had direct negative effects on Rs and NEP, respectively (Fig. 5d and Supplementary Fig. S30). Increasing temperature and extreme precipitation are projected because of climate change, which are associated with a decrease in soil DOM (Fig. 5d). However, the composition of DOM (Fluorescence and MW fraction) can be modulated to decrease soil respiration, increase DOM storage and balance net ecosystem productivity. Long-term grassland management (e.g., grassland conservation and restoration) helps to increase plant biomass and soil clay minerals, which can enhance the ratio of plant-derived components to microbial-derived components and thus restrain soil respiration and increase soil carbon sequestration^{62–64}. On the other hand, the combined variables explained 55 and 90% of the variance of cancer incidence and mortality, respectively (Fig. 5e). Statistically, the DOM quantity (DOM) was significantly associated with cancer mortality, and the fluorescent composition was significantly related to both cancer incidence and mortality (Fig. 5e and Supplementary Fig. S30). This further strengthened the results regarding the potential linkages between the grassland soil DOM stored in the dry season and cancer incidence and mortality. Therefore, our results have important implications for indicating the spatial distribution of endemic cancers, even if DOM is variable under the influences of environmental factors.

Collectively, based on continental-scale sampling, we depict the spatial patterns, drivers, and environment functions of DOM in grassland soils. The geochemistry of the DOM varied across geoclimatic zones and displayed a pronounced South-North divide. Under variable environmental scenarios, LMW proteinaceous fractions and HMW humic-like fractions of soil DOM in grasslands play distinct roles in local ecosystem carbon exchange and imply different human health

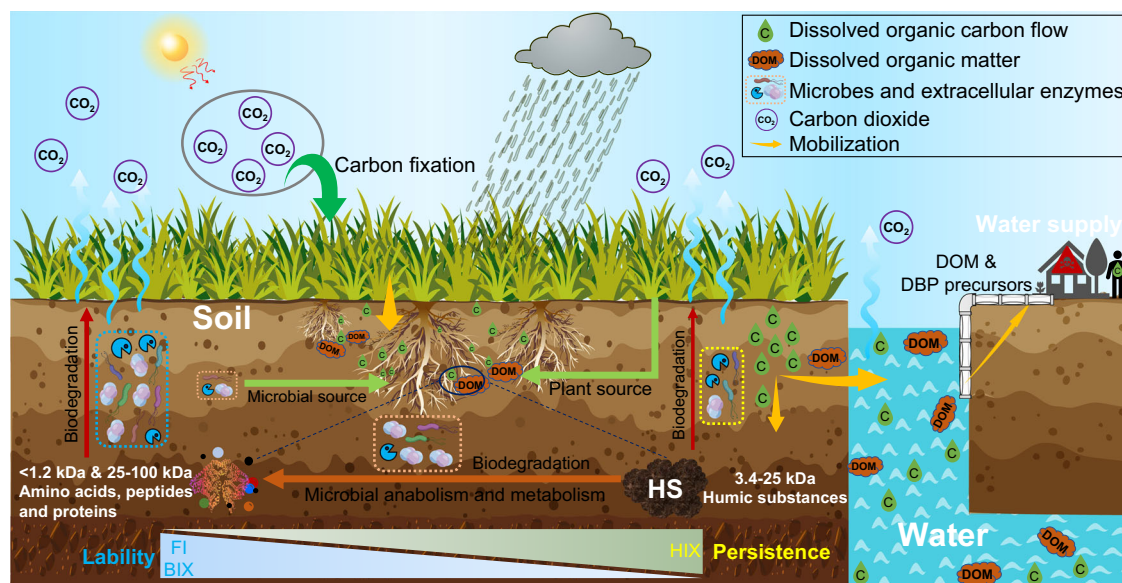


Fig. 6 | Conceptual framework for the key compounds and processes concerning dissolved organic matter (DOM) from surface soil in grassland ecosystems. DOM (mainly in the form of dissolved organic carbon flow) in surface soil (0–20 cm) of grassland was mainly derived from plant sources and undergoes transport and transformation⁷⁷. First, as a portion of soil organic matter, DOM compounds would be processed and mineralized by microbes. Plant-derived humic-like compounds with large molecular mass are more persistent, they are degraded much slower and contribute more to carbon sink⁷⁸. Proteinaceous compounds with low or very high molecular weight, mainly derived from microbial

activities (e.g., processing of humic-like compounds), are readily accessible to microbes and can lead to greater carbon emissions. Second, in addition to migration downwards, DOM can be delivered into aquatic ecosystems via hydrological processes (e.g., run-off or erosion events)⁷⁷. The intruded DOM can alter water chemistry and increase CO₂ outgassing⁷. Moreover, soil DOM in source water supplying drinking water can be ingested by humans directly or in the form of harmful byproducts (e.g., disinfection byproducts), and thus is potentially linked to human health.

conditions in China. Our results provide new insights into the ecosystem processes of grassland soil DOM, which are fundamental to understand the environment functions of terrestrial organic matter in wider processes (Fig. 6). It should be emphasized that the samples in this study were collected in the dry season, and soil DOM may vary seasonally. Our results primarily represent the link between grassland soil DOM in the dry season and the annual ecosystem exchanges and cancer cases, which should be taken into consideration when interpreting the results. Furthermore, microbes play a role in regulating the composition of DOM. Therefore, to fully establish the connections between soil DOM throughout the year and the environmental drivers and functions, observations in other seasons and biotic factors should be considered in further studies.

From a sustainability viewpoint, enriching organic nutrients to foster plant growth is a natural and anticipated aspect of the grassland ecosystems. Although some soil DOM compounds may be related to undesirable environmental impacts, it should be clarified that the complete geochemical cycling of DOM in soils can benefit soil health and environmental stability. On this basis, the environmental functions based on soil DOM geochemistry should be utilized for ecosystem and health benefits: (i) To increase the carbon sequestration potential of grassland soils, reducing overgrazing, restoring vegetation, and decreasing land use intensity of natural grassland are required⁶⁴. These management practices contribute to increasing the input of plant-derived compounds to soils, stabilizing the humic-like components in soils, and mitigating DOM loss caused by biodegradation and erosion⁶³, particularly the increasing loss of organic carbon from grassland soil is projected in the context of climate change^{64,65}. (ii) Given the connections between soil, source water, drinking water, and human health (Fig. 6), the potential adverse effects related to soil DOM in grasslands and broader ecosystems and processes should be minimized via ecological management or engineering measures. For example, in agricultural production, stabilizing the organic fertilizers in soils is necessary to suppress the loss of mobile compounds and

thus reduce their migration into aquatic environments. Further, targeted treatment processes should be applied to remove the associated harmful compounds and prevent the formation of undesirable derivatives (e.g., disinfection byproducts) during drinking water production¹⁶. (iii) In addition, the linkages between DOM and cancer cases could be a crucial roadmap for further studies to fill the knowledge gaps in the identification, prediction, and indication of these endemic cancers of unknown etiology.

Methods

Study sites and soil sampling

Soil samples were collected in May 2021 from 89 geographically dispersed sites of natural grasslands in 30 provinces of China (excluding Hainan, Hongkong, Macao, and Taiwan) (Supplementary Fig. S1). Since we aim to investigate the geochemical features and environmental functions of dissolved organic compounds present in grassland soils, the samples were collected prior to the wet season to avoid the substantial loss of mobile compounds from soils during rainfall and erosion events. The sample areas covered the major Chinese climate, soil and landscape conditions with diverse grassland communities to provide a continental-scale inference for our results. Sample sites were determined based on the spatial distribution of China's grassland. Soil samples were collected from the grasslands undisturbed or less disturbed by direct human activities based on historical data, field surveys, and inquiries with local residents. In each sample site, five soil cores were collected and mixed homogeneously as a composite soil sample. For the regions with a large grassland area, soil cores were collected at a 15 m interval along the diagonal of a 50 m × 50 m square plot in the grassland. For those regions with a small grassland area, soil cores were collected from five dispersed grasslands in each sample site. DOM were extracted and analyzed based on the composite soil sample that represents the value for each sample site. We focused on the topsoil DOM in this study, as organic carbon in the surface soil accounts for more than 40% of that in the 0–1 m soil depth in

grasslands and is more susceptible to mobilization and transformation⁶⁶. Thus, soil samples at 0–20 cm depth were collected, followed by the removal of soil fauna, stones, and plant litter. All soil samples were air-dried, ground, and passed through a 2 mm sieve.

Regions of China and position of sites

The study area was divided into seven regions to compare the spatial difference of soil DOM (Supplementary Fig. S2), including the east coast (EC), middle of Yellow River (MYR), northeast (NE), northwest (NW), south coast (SC), South-North Water Transfer (STN), and southwest (SW) regions. The relative position of sample sites was reflected by the South-North and West-East orientation, and was calculated based on the latitude and longitude coordinates as follows:

$$\text{West-East orientation} = \frac{135^{\circ}05' - X}{135^{\circ}05' - 73^{\circ}33'} \quad (1)$$

$$\text{South-North orientation} = \frac{53^{\circ}33' - Y}{53^{\circ}33' - 3^{\circ}51'} \quad (2)$$

where X and Y are the longitude and latitude of the sample site, respectively; $53^{\circ}33'$ and $3^{\circ}51'$ are the latitudes of the northernmost and southernmost boundaries of China, respectively; and $135^{\circ}05'$ and $73^{\circ}33'$ are the longitudes of the easternmost and westernmost boundaries of China, respectively.

Chemical analysis of water extracts

To avoid the effects of spatial and temporal differences in moisture between soils and to compare the DOM geochemistry in different grassland ecosystems⁶⁷, DOM was extracted from air-dried soils, and thus, the water-extractable organic matter was used to represent soil DOM in this study. Specifically, soil DOM was extracted by shaking (220 rpm and 25 °C) 50 g soil in 500 ml Milli-Q water for 24 h. The soil suspension was thereafter centrifuged at 10,000× g and filtered through pre-combusted glass fiber filters (0.7 μ m, GF/F, Whatman). The water samples were stored in a glass vial in a refrigerator (<4 °C) for subsequent analysis.

The pH, electrical conductivity (ED), and redox potential (Eh) of the water extracts were determined by pH, ED (METTLER TOLEDO, Switzerland), and Eh (501, REX, China) meters, respectively. To reduce the electromagnetic interference in laboratory settings, the redox potential of samples was measured in a fixed place, away from wires and other electrical appliances⁶⁸. All the electrodes were washed with concentrated nitric acid and recalibrated before testing and after each ten measurements. Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) were analyzed by a TOC-L CSH/CSN (Shimadzu, Japan) analyzer. Inorganic nitrogen, including ammonium (NH_4^+) and nitrate and nitrite (NO_x^-), was measured by a continuous flow analyzer (SAN++, Skalar, Holland), and dissolved organic nitrogen (DON) was calculated by subtracting inorganic nitrogen from TDN.

Samples were chlorinated to examine the disinfection byproduct formation potential (DBP-FP) and chlorine reactivity (SDBP) of DOM. Samples were diluted to DOC ~3 mg·L⁻¹ and buffered to pH 8.0 with 3 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ solution, followed by addition of NaOCl solution at a dose of $[\text{Cl}_2] = 3 \times [\text{DOC}] + 7.6 \times [\text{NH}_3]$ ⁶⁹. Then, incubation was conducted at 25 °C in the dark with no headspace for chlorination reaction⁷⁰. The residual chlorine was quenched by 0.5 M ascorbic acid after 72 h. Disinfection byproducts (DBPs) were extracted and analyzed by gas chromatography with an electron capture detector (Clarus 590, PerkinElmer, USA) based on the Method 551.1 and 552.3 from US EPA (details are described in Supplementary Text S2)^{71,72}. After chlorination of the DOM solution, trihalomethanes (THMs) and haloacetic acids (HAAs) were the major detectable DBPs. The specific compounds of DBPs are listed in Supplementary Table S3. The DBP formation potential (DBP-FP; mg·kg-soil⁻¹), representing the

abundance of DBP precursors in soil, was calculated as the ratio of DBP content formed by water extract to the mass of soil, while the chlorine reactivity of DOM was denoted as the specific DBP formation potential (SDBP; mg·g-DOC⁻¹), which was calculated by normalizing DBP concentration to the DOC concentration.

DOM characterization

UV-visible absorbance was analyzed by a UV-visible spectrophotometer (UV-2600, Shimadzu, Japan). A three-dimensional excitation-emission matrix (3D-EEM) of DOM was obtained by a fluorescence spectrophotometer (F-4600, HITACHI, Japan). The UV absorbance at 254 nm (A254), humification index (HIX), biological index (BIX), and fluorescence index (FI) were calculated to show DOM characteristics. A254 and total fluorescence intensity were applied to represent the quantity of chromophoric (CDOM) and fluorophoric (FDOM) DOM, respectively. Greater HIX reflects DOM with higher molecular weight and aromaticity, while greater BIX indicates recently produced DOM with higher freshness⁷³. FI reflects whether the DOM was derived from allochthonous sources (-1.2, plant and soil organic matter) or autochthonous sources (-1.8, bacteria and algae byproducts)⁷³. To further quantify and characterize the composition of fluorescent DOM, parallel factor analysis (PARAFAC) was performed with the DOMFluor toolbox in MATLAB. A four-component model was constructed and validated in this study (Supplementary Table S1), which is comparable to the fluorophores in 63 independent studies in the OpenFluor database (Tucker Congruence Coefficient >0.95). Detailed information regarding the processing of EEM data, calculation of the optical index, and PARAFAC modeling are provided in Supplementary Text S3.

High-performance size exclusion chromatography (HPSEC), employing an HPLC system (Waters, United States) and a UV detector (Phenomenex, UK), was used to determine the apparent molecular weight (MW) distribution of DOM. To characterize the MW composition of DOM, the apparent MW was divided into four MW fractions based on the peak position: very high (VHMW, 25–100 kDa), high (HMW, 3.4–25 kDa), medium (MMW, 1.2–3.4 kDa) and low (LMW, <1.2 kDa) molecular weight fractions. The relative abundance of each fraction was calculated as the ratio of the intensity integral of each MW interval to the total intensity of the four MW fractions.

The molecular composition of fourteen selected samples were further analyzed using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS). Prior to analysis, solid phase extraction (SPE) was conducted using Bond Elute PPL cartridges (Agilent; 200 mg, 3 mL⁻¹). Then, a 15 T Bruker Solarix FT-ICR MS coupled to an electrospray ionization (ESI) interface was applied to analyze the samples. Peaks with signal/noise (S/N) ratio greater than four were assigned molecular formulae following the elemental combination of $\text{C}_{1-100}\text{H}_{1-250}\text{O}_{1-100}\text{N}_{0-3}\text{S}_{0-1}$. Detailed information of SPE processes, FT-ICR MS analysis, and the results are described in Supplementary Text S4 and Table S4.

DOM incubation

To avoid the influence of other complex processes in soils on DOM biodegradation, DOM was extracted from the soils and then incubated^{5,74,75}. Fourteen DOM samples (with seven samples from both the northern and southern China regions, respectively) were incubated at 20 °C in the dark after the addition of microbial inoculum³⁶. The inoculum was prepared at first: (1) 8 g sieved (2 mm) fresh soils from each of the 14 soils and 500 mL Milli-Q water were mixed and incubated at 25 °C in the dark; (2) After 2 weeks, the upper layer suspension was collected and used as the microbial inoculum. Subsequently, the inoculum and DOM solution (at a ratio of 1:100) were added into a pre-combusted 3-L glass bottle filled with 1 kg cleaned and pre-combusted quartz sand for microbial attachment. The C:N:P stoichiometry was amended to 106:16:1 by KNO_3 and KH_2PO_4 to avoid possible nutrient limitation. All bottles were capped by parafilm, and solutions were

stirred twice daily and for 2 min each time. After 31 days, DOM samples were analyzed and the biodegradation rate (% of initial concentration) were calculated to show the biodegradability of DOM. Furthermore, the one-phase association and two-phase association models were fitted in GraphPad Prism 9 to quantify the biodegradation rate constant (d^{-1}) and the labile and stable fractions (%) of DOM (including DOC, CDOM, and FDOM)³⁶, respectively.

Data sources

Environmental factors, including mean annual temperature (MAT), mean annual precipitation (MAP), potential evapotranspiration (PET), surface solar radiation (Srad), leaf area index (LAI), gross primary production (GPP), net primary production (NPP), edaphic variables (including soil moisture, 'SM'; soil clay fraction, 'Clay'; soil pH; cation exchange capacity, 'CEC'; soil exchangeable sodium, potassium and calcium, 'Na', 'K' and 'Ca'; soil organic carbon, 'SOC'; soil total nitrogen, 'TN'; soil total phosphorus, 'TP'), net ecosystem productivity (NEP), and soil respiration (Rs) were collected to examine their associations with the quantity and quality of grassland soil DOM. Details on the environmental factor datasets are provided in Supplementary Text S5 and Table S5. All the datasets were extracted to site-specific values using ArcGIS (v 10.6), and data with several years were averaged to the mean annual values. The cancer incidence and mortality data at the county scale were obtained from 2019 China Cancer Registry Annual Report⁷⁶. To assess the association between grassland soil DOM and local drinking water quality, the total organic carbon (TOC) concentration of tap water for each province of China was collected from a previous study¹⁶.

Statistical analysis

The differences between DOM components and biodegradability were assessed by the Wilcoxon test. Spearman correlation was used to examine the univariate relationships between DOM geochemistry and environmental factors, and between optical properties, molecular composition, and biodegradability. Regression analysis was employed to test the relationships between DOM geochemistry and the position of sites, and between DOM and ecosystem carbon and cancer cases using SPSS 18.0 (IBM, USA). Variation partitioning analysis (VPA) (R package vegan) was conducted to quantify the explanation of size fraction and fluorescence composition to DOM quantity. Principal component analysis (PCA) was conducted using R package vegan, with permutational multivariate analysis of variance (PERMANOVA) applied to examine the group significance. Random forest regression (RF) analyses (R package randomForest) were conducted to explore the key predictors for DOM quantity and quality. Partial least-squares path model (PLS-PM) analysis (R package pls) was applied to reveal the control-response mechanism between environmental factors, DOM, and its environmental functions. The final model was determined from all constructed models based on the goodness of fit (GoF) index that indicates the model performance.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The DOM data supporting the findings of this study are provided within the article and Supplementary Information. The map data (<https://doi.org/10.12078/2023010103>) were obtained from the Resource and Environment Science Data Platform of China (<https://www.resdc.cn/Default.aspx>). The MAT, MAP, and PET data were obtained from the CRU TS v. 4.05 dataset (https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.05/). The Srad data were obtained from the ERA5-Land dataset (<https://www.ecmwf.int/en/era5-land>). The LAI and GPP data were obtained from GLASS products (<http://www.glass.umd.edu>).

The NPP data were obtained from MODIS data (<https://modis.gsfc.nasa.gov>). Soil parameters were obtained from GSDE (<http://globalchange.bnu.edu.cn/research/soilw#download>). The NEP (<https://doi.org/10.12041/geodata.468072.ver1.db>) and Rs (<https://doi.org/10.12041/geodata.3103258.ver1.db>) data were obtained from National Earth System Science Data Center of China (<http://www.geodata.cn>). The cancer incidence and mortality data were obtained from the 2019 China Cancer Registry Annual Report⁷⁶. The TOC concentration of tap water was obtained from ref. 16.

Code availability

The code for random forest regression analyses and partial least-squares path model analysis is available at <https://doi.org/10.6084/m9.figshare.26103484.v1>.

References

- Regnier, P., Resplandy, L., Najjar, R. G. & Ciais, P. The land-to-ocean loops of the global carbon cycle. *Nature* **603**, 401–410 (2022).
- Battin, T. J. et al. The boundless carbon cycle. *Nat. Geosci.* **2**, 598–600 (2009).
- Chen, S. et al. Legacy effects of sorption determine the formation efficiency of mineral-associated soil organic matter. *Environ. Sci. Technol.* **56**, 2044–2053 (2022).
- Ward, C. P. & Cory, R. M. Complete and partial photo-oxidation of dissolved organic matter draining permafrost soils. *Environ. Sci. Technol.* **50**, 3545–3553 (2016).
- Liu, H. et al. Effects of natural vegetation restoration on dissolved organic matter (DOM) biodegradability and its temperature sensitivity. *Water Res.* **191**, 116792 (2021).
- Cole, J. J. et al. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**, 172–185 (2007).
- Li, Y. et al. Changes in water chemistry associated with rainstorm events increase carbon emissions from the inflowing river mouth of a major drinking water reservoir. *Environ. Sci. Technol.* **56**, 16494–16505 (2022).
- Tiwari, T., Sponseller, R. A. & Laudon, H. The emerging role of drought as a regulator of dissolved organic carbon in boreal landscapes. *Nat. Commun.* **13**, 5125 (2022).
- Zeng, X. et al. Recognizing the groundwater related to chronic kidney disease of unknown etiology by humic-like organic matter. *npj Clean. Water* **5**, 8 (2022).
- Makehelwala, M., Wei, Y., Weragoda, S. K., Weerasooriya, R. & Zheng, L. Characterization of dissolved organic carbon in shallow groundwater of chronic kidney disease affected regions in Sri Lanka. *Sci. Total Environ.* **660**, 865–875 (2019).
- Narita, H., Funamizu, N., Takakuwa, T. & Kunimoto, M. Role of hydrophilic organic matter on developing toxicity in decay process of activated sludge. *Water Sci. Technol.* **52**, 63–70 (2005).
- Zeng, X. et al. Molecular linkage of dissolved organic matter in groundwater with prevalence of chronic kidney disease with unknown etiology. *Exposure Health* **15**, 489–503 (2023).
- Huang, T.-S., Lu, F.-J. & Tsai, C.-W. Tissue distribution of absorbed humic acids. *Environ. Geochem. Health* **17**, 1–4 (1995).
- Hua, G., Kim, J. & Reckhow, D. A. Disinfection byproduct formation from lignin precursors. *Water Res.* **63**, 285–295 (2014).
- Richardson, S. D., Plewa, M. J., Wagner, E. D., Schoeny, R. & DeMarini, D. M. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutat. Res.* **636**, 178–242 (2007).
- Liu, M. J. et al. Spatial assessment of tap-water safety in China. *Nat. Sustain.* **5**, 689–698 (2022).
- Jansson, M., Hickler, T., Jonsson, A. & Karlsson, J. Links between terrestrial primary production and bacterial production and

- respiration in lakes in a climate gradient in subarctic Sweden. *Ecosystems* **11**, 367–376 (2008).
18. Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S. & Vivanco, J. M. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu. Rev. Plant Biol.* **57**, 233–266 (2006).
 19. Malik, A. A. et al. Land use driven change in soil pH affects microbial carbon cycling processes. *Nat. Commun.* **9**, 3591 (2018).
 20. Zhu, M. et al. Effects of moisture and salinity on soil dissolved organic matter and ecological risk of coastal wetland. *Environ. Res.* **187**, 109659 (2020).
 21. Tong, H., Simpson, A. J., Paul, E. A. & Simpson, M. J. Land-use change and environmental properties alter the quantity and molecular composition of soil-derived dissolved organic matter. *ACS Earth Space Chem.* **5**, 1395–1406 (2021).
 22. Bahureksa, W. et al. Soil organic matter characterization by Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS): a critical review of sample preparation, analysis, and data interpretation. *Environ. Sci. Technol.* **55**, 9637–9656 (2021).
 23. White, R. P., Murray, S. & Rohweder, M. *Pilot Analysis of Global Ecosystems: Grassland Ecosystems* (World Resources Institute, 2000).
 24. Guo, Z. et al. Soil dissolved organic carbon in terrestrial ecosystems: global budget, spatial distribution and controls. *Glob. Ecol. Biogeogr.* **29**, 2159–2175 (2020).
 25. Ojeda, A. S., Ford, S. D., Gallucci, R. M., Ilnat, M. A. & Philp, R. P. Geochemical characterization and renal cell toxicity of water-soluble extracts from U.S. Gulf Coast lignite. *Environ. Geochem. Health* **41**, 1037–1053 (2019).
 26. Bunnell, J. E., Tatu, C. A., Lerch, H. E., Orem, W. H. & Pavlovic, N. Evaluating nephrotoxicity of high-molecular-weight organic compounds in drinking water from lignite aquifers. *J. Toxicol. Environ. Health A* **70**, 2089–2091 (2007).
 27. Chen, H. et al. Patterns and drivers of the degradability of dissolved organic matter in dryland soils on the Tibetan Plateau. *J. Appl. Ecol.* **59**, 884–894 (2022).
 28. Zhu, E. et al. Leaching of organic carbon from grassland soils under anaerobiosis. *Soil Biol. Biochem.* **141**, 107684 (2020).
 29. Yang, T. et al. Soil fungal diversity in natural grasslands of the Tibetan Plateau: associations with plant diversity and productivity. *N. Phytol.* **215**, 756–765 (2017).
 30. Li, X.-M. et al. Molecular chemodiversity of dissolved organic matter in paddy soils. *Environ. Sci. Technol.* **52**, 963–971 (2018).
 31. Luo, J. et al. Lake chemodiversity driven by natural and anthropogenic factors. *Environ. Sci. Technol.* **56**, 5910–5919 (2022).
 32. Panettieri, M. et al. Grassland-cropland rotation cycles in crop-livestock farming systems regulate priming effect potential in soils through modulation of microbial communities, composition of soil organic matter and abiotic soil properties. *Agric. Ecosyst. Environ.* **299**, 106973 (2020).
 33. Kida, M. et al. Origin, distributions, and environmental significance of ubiquitous humic-like fluorophores in Antarctic lakes and streams. *Water Res.* **163**, 114901 (2019).
 34. Yamashita, Y., Kojima, D., Yoshida, N. & Shibata, H. Relationships between dissolved black carbon and dissolved organic matter in streams. *Chemosphere* **271**, 129824 (2021).
 35. Ding, Y. et al. Chemodiversity of soil dissolved organic matter. *Environ. Sci. Technol.* **54**, 6174–6184 (2020).
 36. Zhou, P. et al. Divergent fate and roles of dissolved organic matter from spatially varied grassland soils in China during long-term biogeochemical processes. *Environ. Sci. Technol.* **58**, 1164–1176 (2024).
 37. D’Andrilli, J., Silverman, V., Buckley, S. & Rosario-Ortiz, F. L. Inferring ecosystem function from dissolved organic matter optical properties: a critical review. *Environ. Sci. Technol.* **56**, 11146–11161 (2022).
 38. Liu, C. et al. Molecular insight into the vertical migration and degradation of dissolved organic matter in riparian soil profiles. *Environ. Res.* **245**, 118013 (2024).
 39. Li, W., Jia, X., Li, M. & Wu, H. Insight into the vertical characteristics of dissolved organic matter in 5-m soil profiles under different land-use types on the Loess Plateau. *Sci. Total Environ.* **692**, 613–621 (2019).
 40. Dai, G. et al. Plant-derived lipids play a crucial role in forest soil carbon accumulation. *Soil Biol. Biochem.* **168**, 108645 (2022).
 41. Kaiser, K. & Kalbitz, K. Cycling downwards - dissolved organic matter in soils. *Soil Biol. Biochem.* **52**, 29–32 (2012).
 42. Thevenot, M., Dignac, M.-F. & Rumpel, C. Fate of lignins in soils: a review. *Soil Biol. Biochem.* **42**, 1200–1211 (2010).
 43. Hao, R. et al. Use of three-dimensional excitation and emission matrix fluorescence spectroscopy for predicting the disinfection by-product formation potential of reclaimed water. *Water Res.* **46**, 5765–5776 (2012).
 44. Hosen, J. D., McDonough, O. T., Febria, C. M. & Palmer, M. A. Dissolved organic matter quality and bioavailability changes across an urbanization gradient in headwater streams. *Environ. Sci. Technol.* **48**, 7817–7824 (2014).
 45. Immerzeel, W. W. et al. Importance and vulnerability of the world’s water towers. *Nature* **577**, 364–369 (2020).
 46. Huang, N., Wang, L., Zhang, Y., Gao, S. & Niu, Z. Estimating the net ecosystem exchange at global FLUXNET sites using a random forest model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 9826–9836 (2021).
 47. He, H. et al. Optical characteristics and cytotoxicity of dissolved organic matter in the effluent and sludge from typical sewage treatment processes. *Sci. Total Environ.* **725**, 138381 (2020).
 48. Lamm, S. H., Boroje, I. J., Ferdosi, H. & Ahn, J. A review of low-dose arsenic risks and human cancers. *Toxicology* **456**, 152768 (2021).
 49. Liu, C.-W., Lai, C.-C., Chen, Y.-Y. & Lu, K.-L. Hydrogeochemical and mineralogical investigations of arsenic- and humic substance-enriched aquifers. *J. Hydrol.* **498**, 59–75 (2013).
 50. Khan, S. et al. Raman spectroscopy combined with principal component analysis for screening nasopharyngeal cancer in human blood sera. *Appl. Spectrosc.* **71**, 2497–2503 (2017).
 51. Souissi, S. et al. Indoleamine 2,3-dioxygenase gene expression and kynurenine to tryptophan ratio correlation with nasopharyngeal carcinoma progression and survival. *Immun. Inflamm. Dis.* **10**, e690 (2022).
 52. Lin, B., Bergholt, M. S., Lau, D. P. & Huang, Z. Diagnosis of early stage nasopharyngeal carcinoma using ultraviolet autofluorescence excitation-emission matrix spectroscopy and parallel factor analysis. *Analyst* **136**, 3896–3903 (2011).
 53. Žuvela, P. et al. Fiber-optic Raman spectroscopy with nature-inspired genetic algorithms enhances real-time in vivo detection and diagnosis of nasopharyngeal carcinoma. *Anal. Chem.* **91**, 8101–8108 (2019).
 54. Burch, J. B., Everson, T. M., Seth, R. K., Wirth, M. D. & Chatterjee, S. Trihalomethane exposure and biomonitoring for the liver injury indicator, alanine aminotransferase, in the United States population (NHANES 1999–2006). *Sci. Total Environ.* **521–522**, 226–234 (2015).
 55. Koivusalo, M. et al. Drinking water mutagenicity and gastrointestinal and urinary tract cancers: an ecological study in Finland. *Am. J. Public Health* **84**, 1223–1228 (1994).
 56. Koivusalo, M., Vartiainen, T., Hakulinen, T., Pukkala, E. & Jaakkola, J. J. K. Drinking water mutagenicity and leukemia, lymphomas, and cancers of the liver, pancreas, and soft tissue. *Arch. Environ. Health* **50**, 269–276 (1995).
 57. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 109 (2020).

58. Jiang, H., Lu, N., Qin, J. & Yao, L. Hourly 5-km surface total and diffuse solar radiation in China, 2007–2018. *Sci. Data* **7**, 311 (2020).
59. Wang, J. et al. New global MuSyQ GPP/NPP remote sensing products from 1981 to 2018. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 5596–5612 (2021).
60. Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).
61. Ma, Y.-J. et al. Water loss by evaporation from China's South-North Water Transfer Project. *Ecol. Eng.* **95**, 206–215 (2016).
62. Chen, S. et al. Plant diversity enhances productivity and soil carbon storage. *Proc. Natl Acad. Sci. USA* **115**, 4027–4032 (2018).
63. Sanderman, J., Maddern, T. & Baldock, J. Similar composition but differential stability of mineral retained organic matter across four classes of clay minerals. *Biogeochemistry* **121**, 409–424 (2014).
64. Liu, L. et al. The grassland carbon cycle: mechanisms, responses to global changes, and potential contribution to carbon neutrality. *Fundam. Res.* **3**, 209–218 (2023).
65. Bai, Y. & Cotrufo, M. F. Grassland soil carbon sequestration: current understanding, challenges, and solutions. *Science* **377**, 603–608 (2022).
66. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
67. Zsolnay, Á. Dissolved organic matter: artefacts, definitions, and functions. *Geoderma* **113**, 187–209 (2003).
68. Husson, O. et al. Practical improvements in soil redox potential (Eh) measurement for characterisation of soil properties. Application for comparison of conventional and conservation agriculture cropping systems. *Anal. Chim. Acta* **906**, 98–109 (2016).
69. Wang, J., Dahlgren, R. A. & Chow, A. T. Controlled burning of forest detritus altering spectroscopic characteristics and chlorine reactivity of dissolved organic matter: effects of temperature and oxygen availability. *Environ. Sci. Technol.* **49**, 14019–14027 (2015).
70. Summers, R. S., Hooper, S. M., Shukairy, H. M., Solarik, G. & Owen, D. Assessing DBP yield: uniform formation conditions. *J. AWWA* **88**, 80–93 (1996).
71. USEPA. Method 551.1. Determination of chlorination disinfection byproducts, chlorinated solvents, and halogenated pesticides/herbicides in drinking water by liquid-liquid extraction and gas chromatography with electron-capture detection (Revision 1.0), Methods for the determination of organic compounds in drinking water (1995).
72. USEPA. Method 552.3. Determination of haloacetic acids and dalapon in drinking water by liquid-liquid microextraction, derivatization, and gas chromatography with electron capture detection (Revision 1.0), Methods for the determination of organic compounds in drinking water (1995).
73. Fellman, J. B., Hood, E. & Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: a review. *Limnol. Oceanogr.* **55**, 2452–2462 (2010).
74. Ohno, T. et al. Molecular composition and biodegradability of soil organic matter: a case study comparing two new England forest types. *Environ. Sci. Technol.* **48**, 7229–7236 (2014).
75. Fellman, J. B., D'Amore, D. V., Hood, E. & Boone, R. D. Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. *Biogeochemistry* **88**, 169–184 (2008).
76. He, J. & Wei, W. *China Cancer Registry Annual Report 2019* (People's Medical Publishing House, 2021).
77. Bianchi Thomas, S. The role of terrestrially derived organic carbon in the coastal ocean: a changing paradigm and the priming effect. *Proc. Natl Acad. Sci. USA* **108**, 19473–19481 (2011).
78. Liang, C., Schimel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2**, 17105 (2017).

Acknowledgements

This research is funded by the National Natural Science Foundation of China (Grants 52370184, W.Y.) and the Beijing Natural Science Foundation (No. JQ21032, W.Y.). The authors would like to thank all those who participated in sample collection.

Author contributions

W.Y., M.E., Y.L., and P.Z. developed the research concepts. P.Z., L.T., S.S., M.S.S., Y.H., and X.C. contributed to collect the data. P.Z., W.Y., M.E., Y.L., N.G., and R.Z. analyzed and interpreted the data. P.Z. wrote the paper and all the authors reviewed the paper and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50745-8>.

Correspondence and requests for materials should be addressed to Yonglong Lu, Menachem Elimelech or Wenzheng Yu.

Peer review information *Nature Communications* thanks Haradhan Kolya, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024