

RESEARCH ARTICLE

Habitat Affinity of Riverine Dissolved Organic Matter Linked to Molecular Traits

Yifan Cui¹ | Ang Hu¹ | James C. Stegen^{2,3} | Jianjun Wang^{1,4} 

¹State Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China | ²Pacific Northwest National Laboratory, Richland, Washington, USA | ³School of the Environment, Washington State University, Pullman, Washington, USA | ⁴University of Chinese Academy of Sciences, Nanjing, China

Correspondence: Jianjun Wang (jjwang@niglas.ac.cn)

Received: 31 October 2025 | **Revised:** 4 January 2026 | **Accepted:** 21 January 2026

Keywords: dissolved organic matter | environmental change | global river | habitat affinity | molecular exchange | molecular trait

ABSTRACT

Fluxes of organic matter across Earth habitats shape ecosystem function and carbon storage. Yet, it is challenging to predict how the molecules are preferentially accumulated in specific habitats, that is, the habitat affinities, due to their diverse characteristics. Here, we develop an indicator of compositional-level habitat affinity for dissolved organic matter (DOM) by quantifying the aggregated affinities of individual molecules that are over- or under-represented relative to a reference habitat. Applying this indicator to 93 paired riverine water-sediment sites, we find that 65.1% of molecules show non-significant habitat affinities for water or sediment and are thermodynamically favorable and susceptible to microbial degradation. Consistently, the habitat affinities of DOM assemblages decrease with lower molecular weights and recalcitrance for waters, and with higher thermodynamic favorability for sediments. These affinities are both reduced by elevated temperatures and anthropogenic pressures, such as nitrogen loading and land-use change. Our projection of habitat affinities of global riverine DOM shows that regions with intense anthropogenic disturbances or warm climates, such as East Asia, Western Europe, and the Amazon, tend to exhibit lower affinities. These findings suggest enhanced molecular homogenization between water and sediment habitats is associated with environmental changes and underscore the importance of habitat affinities for predicting the potential trajectories of organic carbon across ecosystems.

1 | Introduction

Dissolved organic matter (DOM), one of the largest active carbon pools on Earth, fuels biogeochemical cycles and regulates the balance between greenhouse gas emissions and long-term carbon storage (Dittmar et al. 2021; Hansell and Carlson 2014; Tranvik et al. 2009). While DOM spans soils, inland waters, and the ocean (Battin et al. 2009; Catalán et al. 2016; Dittmar et al. 2021; Lehmann and Kleber 2015), rivers play a pivotal role in connecting these pools—transporting an estimated 2.9 Pg C year⁻¹ from terrestrial landscapes toward lakes (~0.6 Pg C year⁻¹) and oceans (~0.9 Pg C year⁻¹), with

the remainder mineralized along the way (Kong et al. 2025; Regnier et al. 2013; Tranvik et al. 2009; Yi et al. 2025). These organic matter fluxes connect spatially segregated ecosystems (or habitats) and shape organic carbon turnover and storage across landscapes (Battin et al. 2009; Freeman et al. 2024; Regnier et al. 2022). Despite the ubiquity of such cross-system connections, most studies focus on the molecular composition and turnover of DOM within a given ecosystem or habitat, such as the water columns or sediments, with limited attention to how molecular exchange across ecosystem boundaries governs the fate of organic matter. This gap limits our understanding of DOM cycling in meta-ecosystems, where matter

Yifan Cui and Ang Hu should be considered joint first authors.

and energy flow between interconnected habitats (Scherer-Lorenzen et al. 2022; Schindler and Scheuerell 2002). In particular, it remains unclear whether individual DOM molecules are preferentially accumulated in specific habitats (i.e., their habitat affinities). Furthermore, how such affinities are shaped by molecular characteristics is poorly understood.

Within a given habitat, DOM composition reflects a dynamic balance between internal processes (such as transformations and losses) and external exchanges with surrounding habitats (Aukes and Schiff 2021; Hu et al. 2025). In river sediments, for instance, some DOM molecules are selectively preserved or degraded depending on environmental conditions and microbial processes (Hu, Choi, et al. 2022; Hu, Jang, et al. 2022). The anoxic or reduced conditions in sediments can promote the potentially long-term persistence of biolabile DOM molecules because high energy costs limit microbial degradation (Hu et al. 2024; LaRowe and Van Cappellen 2011; McDonough et al. 2022). Meanwhile, other molecules may be transported across ecosystem boundaries via largely stochastic processes. Between the sediments and overlying water, DOM is continuously exchanged through physical processes such as sedimentation, hyporheic exchange, mineral adsorption, concentration-gradient diffusion, upwelling, and hydrodynamic perturbation (Creed et al. 2015; Stadler et al. 2023). Such cross-boundary movement can fundamentally alter the fate of DOM due to different biotic and abiotic conditions among habitats.

Yet such molecular exchange within dynamic riverine systems is increasingly shaped by climate change, nutrient loading, and land-use change, which collectively alter the physical and chemical conditions controlling DOM transformation and persistence (Lynch et al. 2019; Wilson and Xenopoulos 2008). Processes such as hyporheic exchange, adsorption–desorption, diffusion, and resuspension regulate the bidirectional flux of DOM molecules between sediments and the water columns (Groeneveld et al. 2020; Riedel et al. 2013). However, a quantitative framework currently remains lacking to resolve the direction and magnitude of individual DOM molecules' preferential accumulation across habitats, nor to assess how these habitat affinities shift in response to environmental change.

Here, we developed a novel indicator to quantify the habitat affinities of DOM for each sample and identify its key environmental drivers, using a riverine molecular dataset generated by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) (Cui et al. 2024; Goldman et al. 2020; Toyoda et al. 2020) (Figure 1a). The dataset includes 273 water and 269 sediment samples from 93 sites along river corridors across three continents of North America, Europe, and Asia (Figure 1a).

This indicator is based on the relative representation of individual molecular formulae (hereafter molecules) between paired water and sediment samples. Two primary procedures were implemented in developing the indicator. First, we calculated the molecule-specific habitat affinity, termed iHaf,

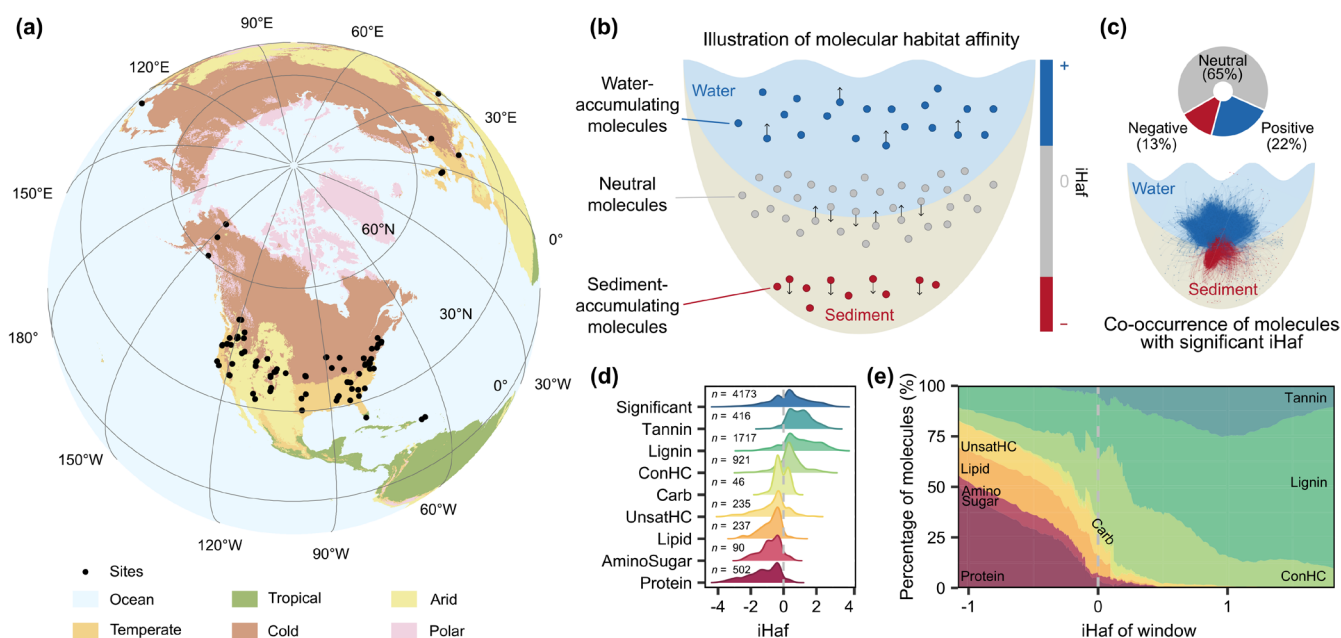


FIGURE 1 | Map of sampling sites and molecular habitat affinity (iHaf) of riverine DOM. (a) Sampling sites across 93 rivers, with the map colored according to the Köppen climate classification (Beck et al. 2018). (b) Conceptual diagram illustrating water-accumulating (blue), sediment-accumulating (red), and neutral molecules (grey). (c) Percentage of molecules showing significant positive, significant negative, and neutral iHaf, along with the co-occurrence network of molecules (shown as nodes) exhibiting significant positive (blue) or negative (red) iHaf. (d) The distribution of statistically significant iHaf values, in aggregate and within each compound class. The compound classes include tannin-like, lignin-like, condensed aromatic-like (ConHC), carbohydrate-like (Carb), unsaturated hydrocarbon-like (UnsatHC), lipid-like, amino sugar-like, and protein-like. n is the number of molecules. (e) The molecular composition at the compound class level (i.e., the percentage of molecules within each compound class) along the continuum of negative to positive iHaf. To construct the iHaf continuum, we used the moving window method to group all molecules into 550 equally sized windows (i.e., 500 molecules per window) according to the magnitude of negative or positive iHaf. The x-axis denotes the lowest absolute iHaf of each window. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

defined as the effect size (Hedges' g) (Hedges 1981; Hedges and Olkin 1985) of each molecule's abundance (i.e., peak intensity) in waters relative to sediments. Based on the effect-size approach, iHaf could provide a statistically robust measure to quantify the magnitude of the habitat affinity for each DOM molecule, enabling direct comparisons across different habitats. Significant positive and negative iHaf values indicate molecules that are preferentially accumulated in waters and sediments, respectively. This could result from the between-habitat imbalance in rates of production, transformation, transport in, and transport out. Non-significant iHaf values closer to zero indicate neutral molecules with higher potential for cross-habitat exchange (Figure 1b). Second, we calculated the abundance-weighted average of iHaf values of individual molecules within each sample to derive a compositional-level indicator (iHaf_{wm}). This upscaling of iHaf to the compositional level further indicates the net representation of water- and sediment-accumulating molecules, providing a more comprehensive framework for assessing the integrated habitat affinity of DOM assemblages. More positive or negative iHaf_{wm} values suggest that a DOM assemblage is increasingly composed of molecules with higher habitat affinities for water and sediment, respectively, while values closer to zero indicate that DOM assemblages are dominated by molecules with greater potential for molecular exchange between habitats, promoting greater homogenization of DOM assemblages across water and sediment habitats.

With this novel approach and a large-scale riverine DOM dataset, we addressed two main questions: (1) Are individual molecules preferentially accumulated across riverine habitats and how are these affinities associated with molecular characteristics? (2) How are the habitat affinities of DOM assemblages influenced by environmental conditions? By explicitly considering the important explanatory variables such as climate and anthropogenic activities, we further provide an exploratory prediction for the spatial patterns of DOM habitat affinities across global rivers. By integrating molecular-level resolution with spatial context, our approach offers new insights into DOM cycling in meta-ecosystems under a changing world.

2 | Materials and Methods

2.1 | Sample Collection

We included a riverine dataset of DOM composition and environmental variables to examine the habitat affinities of DOM and identify its key environmental drivers. This dataset was sourced from the Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) (<https://data.ess-dive.lbl.gov/data>), uploaded by the Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS) consortium (Goldman et al. 2020; Stegen and Goldman 2018; Toyoda et al. 2020). Briefly, river water and sediment samples were collected in triplicate at 97 sites across three continents of North America, Europe, and Asia in July–September 2019. At each site, surface water samples were collected in pre-acidified amber glass vials. Subsequently, sediment samples were collected using a sterilized metal scoop from a depth of 1–3 cm and an area of ~1 m² at depositional zones. Samples were transported

immediately to the Pacific Northwest National Laboratory and then frozen at –20°C. To ensure the paired habitat observations, we selected the sites that contained both water and sediment samples, and ultimately, 273 water and 269 sediment samples from 93 sites were retained for subsequent analyses. These samples cover a wide range of climatic gradients across a large spatial scale (Figure 1a), spanning latitudes from 18.2° to 68.7°N and elevations from –206 to 3013 m, covering five major climatic classifications, including tropical, arid, temperate, cold, and polar. The mean annual temperature (MAT) and mean annual precipitation (MAP) at the sampling sites exhibited broad ranges, with MAT from –10.1°C to 24.1°C and MAP from 168 to 2286 mm. To ensure comprehensive representation of diverse river systems, the sampling covered both main streams and tributaries.

2.2 | FT-ICR MS Analysis of DOM Samples

DOM of river water and sediment samples was pre-processed and solid-phase extracted for FT-ICR MS measurement (Goldman et al. 2020; Toyoda et al. 2020). The detailed procedures for DOM extraction and FT-ICR MS analysis of DOM composition were provided in previous literature (Cui et al. 2024, 2026; Garayburu-Caruso et al. 2020) and briefly described as follows: Thawed water samples and centrifuged sediment samples were filtered using 0.22 μm Millipore filters. Filtrates were acidified to pH2 using 85% phosphoric acid and further solid-phase extracted using Bond Elut PPL cartridges with methanol as eluent, and the extraction efficiency was expected to be over 60% for freshwater samples (Chen et al. 2025; Hu et al. 2024; Roth et al. 2019; Zhan et al. 2024). The non-purgeable organic carbon (NPOC) was analyzed using a Shimadzu combustion carbon analyzer TOC-L CSH/CSN E100V. Organic carbon concentration of DOM was normalized to a standard concentration of 1.5 mg C L⁻¹ across samples before injecting on the mass spectrometry. High-precision mass analysis of DOM extracts was performed via a 12 Tesla Bruker Solarix FT-ICR MS (Bruker, Solarix, Billerica, MA, USA) coupled with a standard electrospray ionization source. Data collection was conducted in negative ionization mode from the mass range of 100–900 m/z , and the assignment of putative chemical formulae followed the same method described previously (Cui et al. 2024). It should be noted that FT-ICR MS is a semi-quantitative approach due to sample matrix effects and differences in ionization efficiencies (Kido Soule et al. 2010). While this technique limitation may affect the quantification of organic molecules, within-molecule comparisons of peak intensities across samples remain meaningful, as supported by recent validation (Kew et al. 2024). Moreover, the consistent methodology and standard DOM concentration were employed across all samples to minimize ionization-related biases.

A total of 11,965 peaks with assigned formulae (hereafter, molecules) were identified for water and sediment DOM samples. Based on the van Krevelen diagrams, assigned molecules were classified into eight compound classes (Hu, Choi, et al. 2022; Kim et al. 2003), including lipid-like (O/C = 0–0.3, H/C = 1.5–2.0), protein-like (O/C = 0.3–0.55, H/C = 1.5–2.2), amino sugar-like (O/C = 0.55–0.67, H/C = 1.5–2.2), carbohydrate-like (Carb; O/C = 0.67–1.1, H/C = 1.5–2.0),

unsaturated hydrocarbon-like (UnsatHC; O/C=0–0.1, H/C=0.7–1.5), lignin-like (O/C=0.1–0.67, H/C=0.7–1.5), tannin-like (O/C=0.67–1.1, H/C=0.5–1.5), and condensed aromatic-like (ConHC; O/C=0–0.67, H/C=0.2–0.7) compounds.

2.3 | DOM Molecular Traits

To evaluate the chemical characteristics of DOM molecules, we calculated 12 molecular traits, including mass, the number of carbon (C) atoms, H/C ratio, O/C ratio, N/C ratio, P/C ratio, S/C ratio, the modified aromaticity index (AI_{mod}) (Koch and Dittmar 2006), double bond equivalents (DBE) (Koch and Dittmar 2006), DBE minus oxygen (DBE_{O}) (Koch and Dittmar 2006), Gibbs free energy (GFE) (LaRowe and Van Cappellen 2011), and carbon use efficiency (Y_{met}) (Song et al. 2020). All traits for each molecule were calculated using the R package iDOM V0.1.0 (Meng et al. 2025). Higher values of H/C ratio and lower values of AI_{mod} and DBE denote lower recalcitrance (D'Andrilli et al. 2015; Šantl-Temkiv et al. 2013). GFE indicates the thermodynamic property of the heterotrophic degradation of organic molecules, and lower GFE of molecules represents more thermodynamically favorable for oxidation (LaRowe and Van Cappellen 2011). For each DOM sample, compositional-level molecular traits were calculated using the abundance-weighted method, which was performed through R package FD V1.0-12 (Lavorel et al. 2008). Take the H/C ratio as an example:

$$H/C_{\text{wm}} = \frac{\sum (H/C_i \times I_i)}{\sum I_i} \quad (1)$$

where H/C_{wm} is the abundance-weighted average value of H/C ratio at the compositional level, H/C_i is the H/C ratio of molecule i , and I_i is the relative abundance of molecule i .

2.4 | Environmental Variables

We collected five categories of environmental variables, including hydrography, climate, anthropogenic influences, land cover and soil, and chemistry variables (Table S1). Hydrography, climate, and chemistry variables were directly extracted at the sampling sites. Anthropogenic influences, as well as land cover and soil properties, were extracted using a spatial buffer of 5 km around the sampling sites (Picazo et al. 2020).

Hydrography variables included elevation, river order, and river discharge. Elevation for each sampling site was derived from the Multiple-Error Removed Improved-Terrain Digital Elevation Model with the resolution of ~90 m (Yamazaki et al. 2017). River order was extracted from HydroRIVERS (Lehner and Grill 2013). River discharge was obtained from the Global Reach-Scale A Priori Discharge Estimates (Lin et al. 2019).

Climatic variables included the measured water temperature, 19 bioclimatic variables, ultraviolet radiation, and wind speed.

Water temperature was measured by WHONDRS consortium (Goldman et al. 2020; Toyoda et al. 2020). The bioclimatic variables were obtained from WorldClim 2.0 (~1 km resolution) (Fick and Hijmans 2017). The temperature-related bioclimatic variables were mean annual temperature, mean diurnal range, isothermality, temperature seasonality, maximum temperature of warmest month, minimum temperature of coldest month, temperature annual range, mean temperature of wettest quarter, mean temperature of driest quarter, mean temperature of warmest quarter, and mean temperature of coldest quarter. The precipitation-related bioclimatic variables included mean annual precipitation, precipitation of wettest month, precipitation of driest month, precipitation seasonality, precipitation of wettest quarter, precipitation of driest quarter, precipitation of warmest quarter, and precipitation of coldest quarter. The ultraviolet radiation and wind speed were extracted from the ERA5 monthly average data (0.25° resolution; <https://cds.climate.copernicus.eu>) (Hersbach et al. 2023).

Anthropogenic influences included population density (~1 km resolution; <https://www.worldpop.org/>), human footprint index (0.00989° resolution) (Keys et al. 2021), gross domestic product (5 arc-min resolution) (Kummu et al. 2018), and the nitrogen and phosphorus fertilizer use rates (0.5° resolution) (Lu and Tian 2017).

Land cover and soil properties included 10 types of land cover, three types of soil textures, and soil organic carbon. Land cover data included cropland, forest, grassland, shrubland, wetland, water body, tundra, impervious area, bare land, and snow/ice, which were collected from the Finer Resolution Observation and Monitoring of Global Land Cover database with the resolution of 30 m (Li et al. 2017). Soil organic carbon and soil texture included soil sand, clay, and silt contents, which were obtained from the Harmonized World Soil Database (v1.2) with the resolution of ~1 km (Fischer et al. 2008), accessible via the Food and Agriculture Organization of the United Nations soils portal.

Chemistry variables included water pH, NPOC, water total nitrogen (TN), sediment TN, carbon/nitrogen ratio, and dissolved inorganic carbon, which were measured for each sample as described in Toyoda et al. (2020) and Goldman et al. (2020).

2.5 | Development of the Indicator of DOM Habitat Affinity

We developed a novel indicator to quantify the habitat affinity of DOM for each sample. This indicator reflects the aggregated affinities of individual DOM molecules that are over- or under-represented relative to a reference habitat. Two primary procedures were implemented in developing the indicator: (1) calculating the direction and magnitude of molecule-specific habitat affinity (termed iHaf), and (2) producing the compositional-level habitat affinity for each sample (termed $iHaf_{\text{wm}}$). Using the relative abundance of individual DOM molecules across water and sediment samples as input, these procedures produced iHaf values for each molecule and $iHaf_{\text{wm}}$ values for each sample.

First, we calculated the iHaf values for individual DOM molecules. The iHaf was defined as the effect size (Hedges' g) (Hedges 1981; Hedges and Olkin 1985) of each molecule's abundance in waters relative to sediments:

$$\text{iHaf} = \frac{\overline{\text{RA}}_w - \overline{\text{RA}}_s}{s_{\text{pooled}}} \times \left(\frac{\Gamma(\text{df}/2)}{\sqrt{\text{df}/2\Gamma((\text{df}-1)/2)}} \right) \quad (2)$$

$$\text{df} = n_w + n_s - 2 \quad (3)$$

$$s_{\text{pooled}} = \sqrt{\frac{\sum (\text{RA}_w - \overline{\text{RA}}_w)^2 + \sum (\text{RA}_s - \overline{\text{RA}}_s)^2}{n_w + n_s - 2}} \quad (4)$$

where $\overline{\text{RA}}_w$ and $\overline{\text{RA}}_s$ are the mean relative abundance of individual molecules in water and sediment habitats, respectively. Γ is the gamma function. n_w and n_s are the sample numbers of waters and sediments, respectively. s_{pooled} is the pooled standard deviation of the two habitats. The calculation of effect size was performed using the R package metafor V4.8-0 (Viechtbauer 2010). The significance of iHaf values was determined using the Kruskal-Wallis test. Significantly ($p \leq 0.05$) positive and negative iHaf values indicate molecules that are preferentially accumulated in waters and sediments, respectively. Neutral ($p > 0.05$) iHaf values closer to zero indicate low habitat affinities of molecules with higher potential for cross-habitat exchange.

Second, we integrated the iHaf values of all molecules within a DOM sample into iHaf_{wm} to represent the net habitat affinity of the DOM assemblage. The iHaf_{wm} was calculated as the abundance-weighted average of iHaf values of individual molecules within each DOM sample through R package FD V1.0-12 (Lavorel et al. 2008):

$$\text{iHaf}_{\text{wm}} = \frac{\sum (\text{iHaf}_i \times I_i)}{\sum I_i} \quad (5)$$

where iHaf_{*i*} is the iHaf value of molecule *i*, and *I_i* is the relative abundance of molecule *i*. More positive or negative iHaf_{wm} values indicate that a DOM assemblage is increasingly composed of molecules with higher habitat affinities for water and sediment, respectively. iHaf_{wm} values closer to zero indicate that DOM assemblages are dominated by molecules with greater potential for molecular exchange between habitats, promoting greater homogenization of DOM assemblages across water and sediment habitats.

To ensure that the iHaf_{wm} calculation is independent of iHaf calculation, we applied independent datasets for each calculation. That is, the entire input dataset of DOM composition was randomly divided into two independent subsets, as initially described in Hu et al. (2024). One dataset was used to calculate iHaf values, while the remaining dataset was employed for iHaf_{wm} calculation. This approach avoids the relative abundance of individual molecules from being used repeatedly in both iHaf and iHaf_{wm} calculations. We used a split ratio of 80:20 for data separation, a common practice in machine learning and species distribution modeling (Naimi

and Araújo 2016). For each habitat, 80% of the samples were allocated to calculate the iHaf for each molecule, while the remaining 20% were used to calculate the iHaf_{wm} for each sample. To obtain the robust iHaf and iHaf_{wm} indicators, the random data splitting and subsequent indicator calculations were repeatedly performed 999 times. The iHaf and iHaf_{wm} values were then averaged across these iterations for use in subsequent statistical analyses. Notably, iHaf values for individual molecules, calculated using all 273 water samples and 269 sediment samples, were essentially identical to those derived from the averaged results of 999 randomized iterations using 80% of the samples, showing a linear regression slope of 1.

2.6 | Statistical Analyses

2.6.1 | Co-Occurrence Network of Water- and Sediment-Accumulating Molecules

To visualize the co-occurrence of water- and sediment-accumulating molecules, we constructed a co-occurrence network based on molecules with significant ($p \leq 0.05$) iHaf values using the R package igraph V2.2.1 (Csárdi and Nepusz 2006). Briefly, we calculated pairwise Spearman rank correlations ρ among the relative abundance of DOM molecules. The *p* values were adjusted using the false discovery rate (FDR) correction. To avoid spurious correlations, we selected molecules that were detected in over 20% of samples. To ensure the robustness of the co-occurrence network, we considered Spearman's rank correlations ρ that were > 0.65 and significance levels of *p* value < 0.00001 (Delgado-Baquerizo et al. 2018). The network that we constructed contained 2112 nodes and 382,681 edges. Of these, 68.9% of molecules (1456 nodes) were water-accumulating with significant ($p \leq 0.05$) positive iHaf, while 31.1% (656 nodes) were sediment-accumulating with significant ($p \leq 0.05$) negative iHaf. We visualized the network using Gephi software (V0.10.1) (Bastian et al. 2009).

2.6.2 | Molecular Habitat Affinity Associated With Molecular Characteristics

At the molecular level, our focus was on how the DOM composition and chemical characteristics, such as recalcitrance and thermodynamic properties, changed gradually or sharply toward lower habitat affinities of individual molecules. Using the moving window method (Felipe-Lucia et al. 2020; Hu, Jang, et al. 2022), we created a continuum of iHaf and corresponding molecular composition and traits. Briefly, we first calculated iHaf values for all molecules and categorized them into two groups, that is, negative and positive iHaf. The molecules within each group were further sorted along the gradient of habitat affinities from the lowest to highest absolute values of iHaf. Using a window size of 500 molecules and a step of 20 molecules, we created 219 consecutive windows with negative iHaf values (i.e., molecules 1–500, 21–520, ..., 4361–4860), and 331 consecutive windows with positive iHaf values (i.e., molecules 1–500, 21–520, ..., 6601–7100). We regarded the lowest absolute values of iHaf for each of the 550 windows as the continuum of

iHaf, ranging from -1.08 to 1.83 . We then examined the characteristics of molecular composition and traits for each window, including the percentage of molecules within each compound class relative to the total number of assigned peaks, and the average values of molecular traits like mass, H/C ratio, O/C ratio, GFE, AI_{mod} , and DBE. To confirm the robustness of the results, we performed a sensitivity analysis with window sizes of 300 and 700 molecules and steps of 10 and 30 molecules, respectively. Consistent patterns were observed in the characteristics of molecular composition and traits along the continuum of iHaf (Figures S12 and S13).

In addition, we performed the principal component analysis (PCA) to assess how the habitat affinities of DOM molecules are explained by their inherent chemical traits, such as recalcitrance and thermodynamic properties. The PCA was conducted using the R package stats V4.2.1 with 12 molecular traits, including mass, number of C atoms, H/C ratio, O/C ratio, N/C ratio, P/C ratio, S/C ratio, AI_{mod} , DBE, DBE_O , GFE, and Y_{met} . The first two axes of the PCA were used to define a two-dimensional orthogonal space. In the PCA space, all molecules were grouped into nine iHaf intervals, including the neutral ($p > 0.05$) iHaf and significant ($p \leq 0.05$) iHaf ranges in < -2 , $[-2, -1]$, $[-1, -0.5]$, $[-0.5, 0]$, $[0, 0.5]$, $[0.5, 1]$, $[1, 2]$, and ≥ 2 . To examine whether molecular traits converge or diverge across the continuum of iHaf values, we calculated the distances from individual molecules to the centroid of the PCA space within each iHaf interval.

To examine how molecular habitat affinity is associated with ecosystem function, we analyzed the relationship between molecular abundance and sediment respiration rates along the continuum of iHaf from negative or positive to zero, corresponding to a gradient from high to low molecular habitat affinities. For each window, we calculated the Spearman rank correlation coefficient ρ between the total relative abundance of molecules and the sediment respiration rates in the corresponding samples (Cui et al. 2024; Kellerman et al. 2014).

2.6.3 | Variation in iHaf_{wm} Along Molecular Traits and Environmental Gradients

To examine the relationships between compositional-level habitat affinity (i.e., iHaf_{wm}) and molecular traits, we used Pearson correlation, hierarchical clustering, and linear regression models with the R package stats V4.2.1. First, we calculated the Pearson correlation coefficient r between iHaf_{wm} and 12 abundance-weighted traits. Second, we performed the hierarchical cluster analysis of iHaf_{wm} and molecular traits based on the absolute Pearson correlation coefficients to identify the molecular traits that are closely associated with DOM habitat affinity. Third, we conducted linear regression models to visualize the relationships between iHaf_{wm} and molecular traits.

To explore the geographical patterns of habitat affinities for DOM assemblages, we analyzed its variations with latitude and elevation using linear regression models. To identify the potential environmental drivers of compositional-level habitat affinity, we calculated the Pearson correlation coefficients r between iHaf_{wm} and hydrography, climate, anthropogenic influences, land cover and soil, and chemical variables (Table S1). For

non-normally distributed data, the natural-log transformation was applied. We further analyzed the variations of iHaf_{wm} with environmental variables using linear regression models.

2.6.4 | Global Mapping of Habitat Affinities for Riverine DOM Assemblages

Using the machine learning-based random forest model, we generated an exploratory prediction for the habitat affinities of DOM assemblages in global river waters and sediments by including 43 environmental variables related to hydrography, climate, anthropogenic influences, land cover, and soil properties (Table S1). Given that our training dataset is mainly derived from North American river systems, this global extrapolation provides a spatially explicit hypothesis that can be tested and refined through future sampling in underrepresented regions. We used global river networks obtained from the HydroRIVERS dataset (Lehner and Grill 2013), focusing exclusively on rivers with a river order higher than 5. To reduce multicollinearity, predictors with strong correlations (absolute Pearson correlations higher than 0.80) and high variance inflation factors (VIF > 10) were not considered in the model, and 25 environmental variables were retained in the final model (Figure S9).

We performed the random forest model using tenfold cross-validation, splitting the data into 80% for training and the remaining 20% for validation (Wang et al. 2024). The root mean square error (RMSE) and determination coefficient (R^2) were used to evaluate model performance, with the optimal model being selected based on the highest R^2 value. The importance of individual variables in predicting the DOM habitat affinity was quantified by assessing the increase in mean square error, that is, the decline of prediction accuracy. The importance measure for accuracy was calculated for each tree and then averaged over the 2000 trees generated (Elith et al. 2008). The relative importance of each predictor was determined by normalizing its importance against that of the most important predictor and multiplying by the determination coefficient (R^2) of the model. We used the optimal random forest model with minimal RMSE to predict DOM habitat affinities for both waters and sediments across the global rivers. The uncertainty of DOM habitat affinity was quantified as the standard error of individual predictions of 2000 trees in the random forest model (Figure S11). The above analysis was implemented via the R package caret V6.0–94.

3 | Results and Discussion

3.1 | Habitat Affinities of Individual DOM Molecules

Across riverine ecosystems, individual DOM molecules exhibited substantial variation in their habitat affinities (Figure 1d). Approximately one-third of molecules showed significant ($p \leq 0.05$) differences in relative abundance between paired water and sediment samples. Among them, 22.4% ($n = 2682$) molecules were preferentially accumulated in water habitats with significant positive iHaf values, while 12.5% ($n = 1491$) molecules were preferentially accumulated in sediment habitats with significant negative iHaf values (Figure 1c and Figure S1).

The iHaf values ranged from -4.13 to 3.70 , with median values of -0.65 and 1.04 for sediment- and water-accumulating molecules, respectively (Figure 1d). In contrast, the majority of molecules (65.1%, $n=7792$) exhibited non-significant ($p>0.05$) habitat affinities, with iHaf values ranging from -0.59 to 0.34 (Figure 1c and Figure S2). This predominance of non-significant iHaf values indicates that most molecules have a greater potential for cross-habitat exchange between water and sediment.

These disparate habitat affinities were associated with distinct molecular compositions, characterized by specific compound classes (Figure 1d,e and Figure S3). For instance, iHaf distributions of tannin- and lignin-like compounds showed notable skewness or spread toward positive values, which could be water-accumulating molecules, while those of lipid-, amino sugar-, and protein-like compounds were skewed toward negative values, which could be sediment-accumulating molecules (Figure 1d). In contrast, condensed aromatic- and carbohydrate-like compounds showed iHaf values centered around zero, indicating low habitat affinity (Figure 1d). To assess whether the molecular composition changes gradually or sharply across the habitat affinity spectrum, all molecules were grouped into 550 windows along the continuous gradient of iHaf values (see Section 2). The percentage of lignin-like compounds in a molecule window increased as the iHaf of the window became more positive, reaching up to 80.2% in the windows of strong water-accumulating molecules (Figure 1e and Figure S3). Conversely, protein-like compounds increased with more negative iHaf values, reaching up to 51.2% in the windows of strong sediment-accumulating molecules (Figure 1e and Figure S3). The percentage of condensed aromatic-like compounds peaked (up to 55.2%) in windows with iHaf values near zero (Figure 1e and Figure S3). These results indicate that molecules with low habitat affinity are largely composed of condensed aromatic-like compounds, with lignin- and protein-like compounds increasing along the habitat affinity gradients of waters and sediments, respectively.

By integrating ultrahigh-resolution mass spectrometry with our newly developed indicator iHaf, we quantified for the first time both direction (toward water or sediment) and magnitude in the habitat affinities of individual molecules across riverine ecosystems. The concept of habitat preference, commonly used in ecology to describe the probability that an organism selects one habitat over another when alternative choices are offered on an equal basis (Beyer et al. 2010; Johnson 1980), has been widely applied to animals, for example, insects and mammals (Li et al. 2025; Santoiemma et al. 2019), plants (Arianoutsou et al. 2010; Maire et al. 2012), and microbes (Delgado-Baquerizo et al. 2018; Wang et al. 2013). Similarly, the habitat affinity concept can be adapted to DOM, as molecules are selectively preserved or lost in a given habitat due to differential production and degradation rates, analogous to the birth-death dynamics in biological communities (Hu, Jang, et al. 2022).

Notably, water and sediment habitats exhibited distinct molecular-level affinities in both the number of significantly represented molecules and the magnitude of iHaf values. More molecules were preferentially accumulated in waters than in sediments. This asymmetry may result from the influx of terrestrial organic matter into river waters from upstream catchments,

with only a fraction settling into sediments via selective adsorption (Riedel et al. 2016; Sowers et al. 2019). Moreover, the generally stronger water affinities (i.e., higher absolute iHaf values) relative to sediment affinities may reflect the longer residence time and enhanced microbial activity in sediments, which promote the degradation of organic molecules (Danczak et al. 2021; Stegen et al. 2018). These findings support the concept that river water and sediment represent essentially isolated habitats despite significant hydrologic exchange, with substantial influences of watershed processes on surface water molecules but highly localized impacts on sediment molecules (Stegen et al. 2022). Together, these findings highlight the potential of iHaf as an indicator for identifying representative DOM molecules within and across natural habitats, offering a quantitative framework to predict their cross-habitat spatial distributions.

3.2 | Habitat Affinity Associated With Molecular Traits

The contrasting habitat affinities of DOM molecules could be partly explained by their inherent chemical traits, such as recalcitrance and thermodynamic properties (Figure 2a and Figure S4). Overall, water- and sediment-accumulating molecules occupied distinct positions in a two-dimensional orthogonal space of principal component analysis (PCA) using 12 molecular traits (see Section 2). Among these traits, double bond equivalents (DBE) and DBE minus oxygen contributed the most variance for PC1, and Gibbs free energy (GFE) and H/C ratio showed the most contributions for PC2 (Figure S4). We further examined how molecular traits change across the continuum of iHaf values. Molecules with higher habitat affinities (i.e., iHaf <-1 or >1) tended to cluster more tightly in the Van Krevelen space and showed more reduced distances to the centroid in the PCA space (Figure 2b,c), suggesting that higher habitat affinity is associated with greater molecular functional similarity within a given habitat.

Importantly, the traits of molecules with higher habitat affinities differed markedly between habitats (Figure 2d). Compared with water-accumulating molecules, sediment-accumulating molecules were less recalcitrant (characterized by lower molecular weight and higher H/C ratio) and less thermodynamically favorable for oxidation (i.e., higher GFE) (Figure 2d). These trait contrasts reveal divergent selective mechanisms: recalcitrance favors the accumulation and persistence of molecules in waters, while thermodynamic limits govern their accumulation in sediments.

More interestingly, as iHaf in a molecule window increased from negative toward zero along the iHaf continuum, sediment-accumulating molecules exhibited a notable decline in GFE, from 84 to 68 kJ (mol C) $^{-1}$, approaching values similar to those of water-accumulating molecules (Figure 2d). This shift suggests that molecules with low habitat affinities, and thus higher potential for cross-habitat exchange between habitats, tend to be more thermodynamically favorable and susceptible to microbial degradation. The microbial degradation of thermodynamically favorable molecules could prevent preferential accumulation in sediments, causing similar relative abundance of molecules and promoting molecular homogenization between

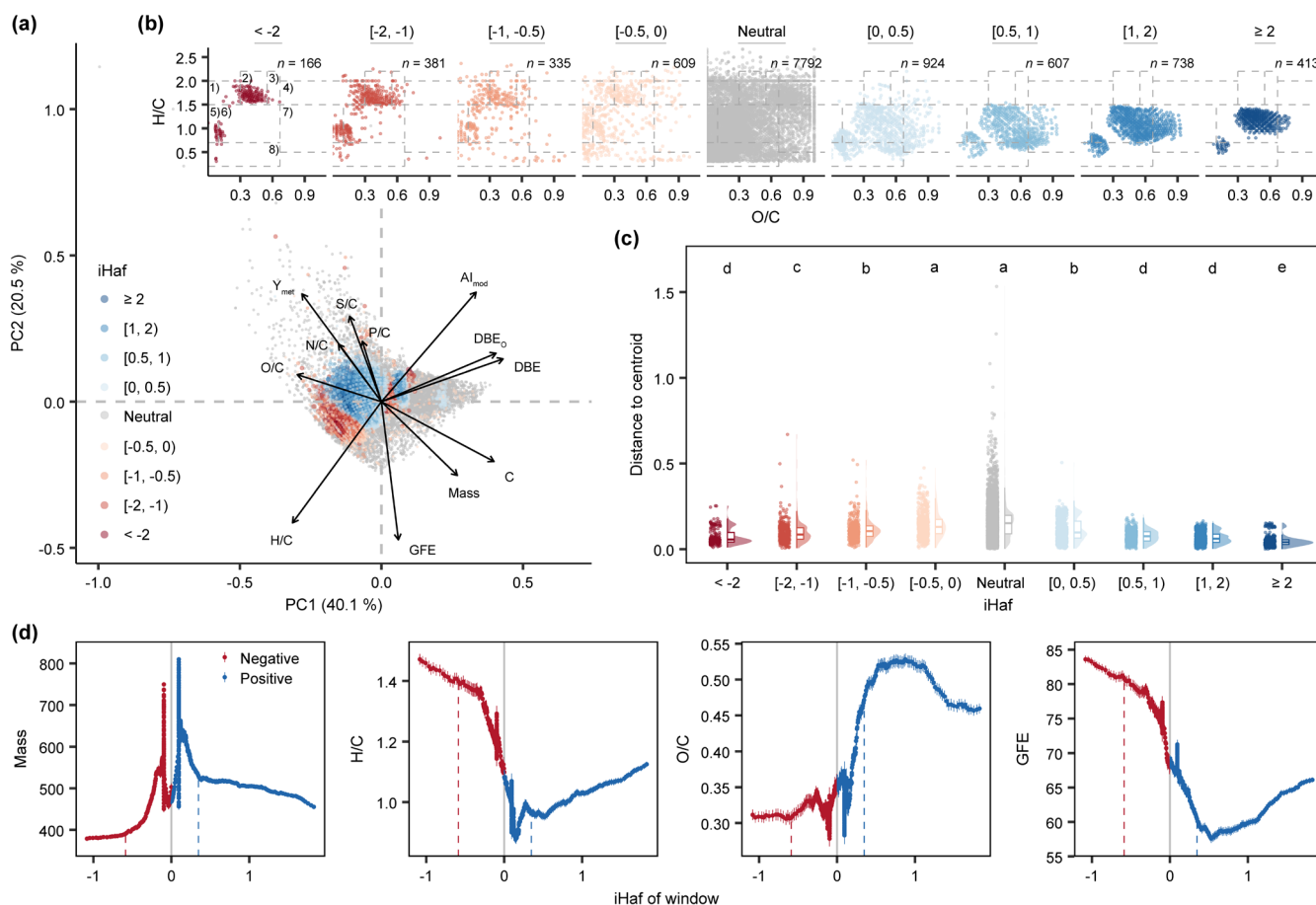


FIGURE 2 | Relationship between molecular habitat affinity (iHaf) and molecular traits. (a) Molecular traits in a two-dimensional orthogonal space of principal component analysis (PCA) used to explain iHaf. Abbreviations of molecular traits are detailed in Table S2. Points are colored with iHaf values of DOM molecules. (b) The distribution of DOM molecules along the continuum of iHaf from negative (red) to positive (blue) displayed in van Krevelen diagrams. We grouped all molecules into nine iHaf intervals, including neutral iHaf and significant iHaf ranges in < -2 , $[-2, -1)$, $[-1, -0.5)$, $[-0.5, 0)$, $[0, 0.5)$, $[0.5, 1)$, $[1, 2)$, and ≥ 2 . n is the number of molecules. The grey dashed lines separating compound classes in van Krevelen diagrams are for visualization only. The compound classes include (1) lipid-like, (2) protein-like, (3) amino sugar-like, (4) carbohydrate-like, (5) unsaturated hydrocarbon-like, (6) lignin-like, (7) tannin-like, and (8) condensed aromatic-like. (c) Distances from individual molecules to the centroid of the PCA space within each iHaf interval. Smaller dots are the values for individual molecules. Boxes indicate interquartile range, and thick horizontal lines within boxes denote median values. Different letters indicate significant ($p < 0.05$) differences based on the Kruskal-Wallis test. (d) Changes in molecular traits along the continuum of iHaf as in Figure 1e. The values of molecular traits are presented as the means \pm SE. Dashed lines in blue and red mark the thresholds of 100% significantly positive and negative iHaf for the molecules in each window, respectively.

water and sediment habitats. Notably, the relative abundance of low-habitat-affinity molecules (iHaf between -0.3 and 0.2) was significantly positively related to sediment respiration rates (Figure S5), causing the preferential removal of labile molecules and the persistence of more recalcitrant ones, as indicated by a reduction in H/C ratio (from 1.47 to 1.10) and an increase in aromaticity index (from 0.19 to 0.40) as iHaf increased from negative toward zero (Figure 2d and Figure S6).

3.3 | Geographical Patterns of Habitat Affinities for DOM Assemblages

iHaf can be upscaled to the sample level using weighted means (iHaf_{wm}), which represent the net representation of water- and sediment-accumulating molecules within a DOM assemblage. In waters, iHaf_{wm} values ranged from 0.24 to 2.30, with a positive mean (\pm SD) of 1.67 (\pm 0.26) (Figure S7). In sediments,

iHaf_{wm} values ranged from -1.40 to 0.93 , with a lower negative mean (\pm SD) of -0.70 (\pm 0.44) (Figure S7). On average, the magnitude of iHaf_{wm} was 2.4 times higher in waters than in sediments (Figure S7).

There were distinct geographical patterns regarding the habitat affinities of DOM assemblages across riverine ecosystems. Specifically, iHaf_{wm} in waters increased significantly with increasing elevation ($r = 0.35$, $p < 0.001$), while iHaf_{wm} in sediments declined significantly toward higher elevation ($r = -0.15$, $p = 0.014$; Figure 3a). Every 100 m increase in elevation was associated with a 0.7% increase in iHaf_{wm} for waters and a 1.2% decrease for sediments, suggesting that DOM habitat affinity tends to be weaker at lower elevations. In contrast, no clear latitudinal pattern was observed in waters and sediments (Figure 3a). Mechanistically, lower elevations are characterized by gentler river gradients, which prolong water residence time and facilitate cross-habitat molecular exchange (Battin et al. 2008;

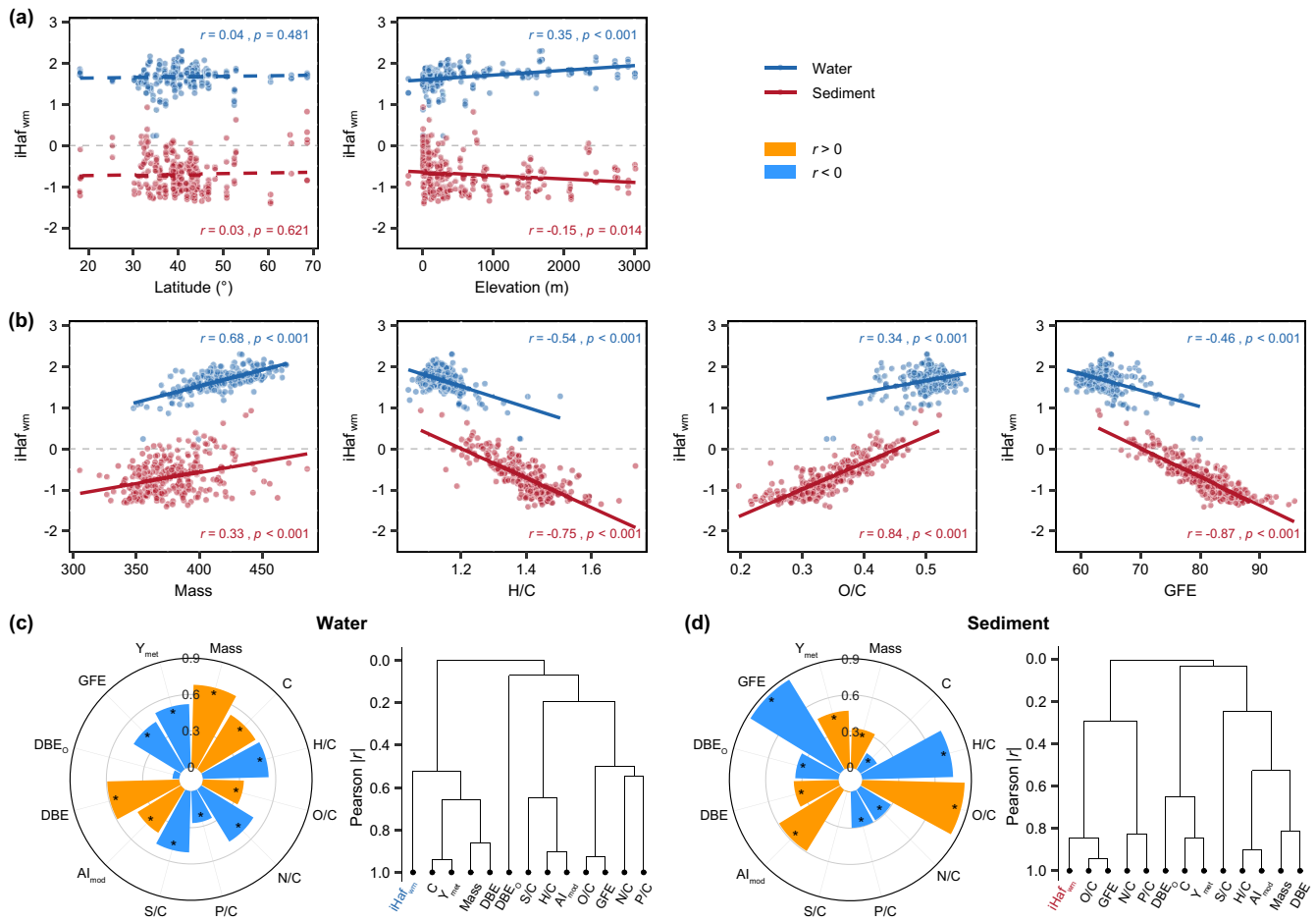


FIGURE 3 | The compositional-level habitat affinity ($iHaf_{wm}$) of DOM along the geographical gradient and its association with molecular characteristics. (a) The habitat affinities of DOM assemblages along latitudinal and elevational gradients in waters (blue) and sediments (red). (b) Variations in habitat affinities of DOM assemblages with abundance-weighted traits, including mass, H/C ratio, O/C ratio, and Gibbs free energy (GFE). Solid and dashed lines in (a) and (b) indicate statistically significant ($p \leq 0.05$) and non-significant ($p > 0.05$) relationships, respectively. (c, d) Relationships between the habitat affinities of DOM assemblages and molecular traits in (c) waters and (d) sediments. Left panels: Pearson correlation coefficient r between $iHaf_{wm}$ and molecular traits. The color of bar indicates the direction of the correlation (positive or negative). The asterisks indicate statistically significant ($p \leq 0.05$) relationships. Right panel: Cluster analysis of $iHaf_{wm}$ and molecular traits base on the absolute Pearson correlation coefficient $|r|$. Abbreviations of molecular traits are detailed in Table S2.

Harvey and Gooseff 2015). Moreover, lower elevations are associated with higher temperatures, which could promote molecular exchange between habitats, a finding further confirmed by subsequent results.

As expected, DOM habitat affinities at the assemblage level were closely linked to recalcitrance and thermodynamic properties. In waters, $iHaf_{wm}$ correlated positively with molecular weight ($r = 0.68, p < 0.001$) and DBE ($r = 0.59, p < 0.001$; Figure 3b,c and Figure S8). In sediments, $iHaf_{wm}$ showed the strongest correlation with GFE ($r = -0.87, p < 0.001$), followed by O/C ratio ($r = 0.84, p < 0.001$) and H/C ratio ($r = -0.75, p < 0.001$; Figure 3b,d). These results indicate that DOM assemblages with $iHaf_{wm}$ values near zero are enriched in smaller molecules with lower recalcitrance in waters, and higher thermodynamic favorability (i.e., decreased in Gibbs free energy) in sediments.

The magnitude of $iHaf_{wm}$ reflects the potential for molecular exchange between habitats, with values near zero indicating greater molecular homogenization between water and sediment

habitats and a higher likelihood of undergoing transformation and driving DOM turnover. Our findings show that DOM assemblages at lower elevations contain more such molecules, suggesting a link between cross-habitat exchange potential and environmental conditions such as climate and anthropogenic activities. Thus, $iHaf$ serves as a straightforward yet powerful metric for assessing DOM habitat affinity in various ecosystems and predicting the fate of DOM under environmental change.

3.4 | Environmental Conditions Are Linked to DOM Habitat Affinity

The habitat affinities of riverine DOM assemblages were significantly influenced by environmental conditions, particularly elevated temperatures and anthropogenic pressures. Specifically, $iHaf_{wm}$ in waters was negatively correlated with water temperature ($r = -0.38, p < 0.001$), while $iHaf_{wm}$ in sediments was positively correlated ($r = 0.15, p = 0.015$) (Figure 4). A 1°C increase in temperature was linked to a decrease in $iHaf_{wm}$ by 0.9% in

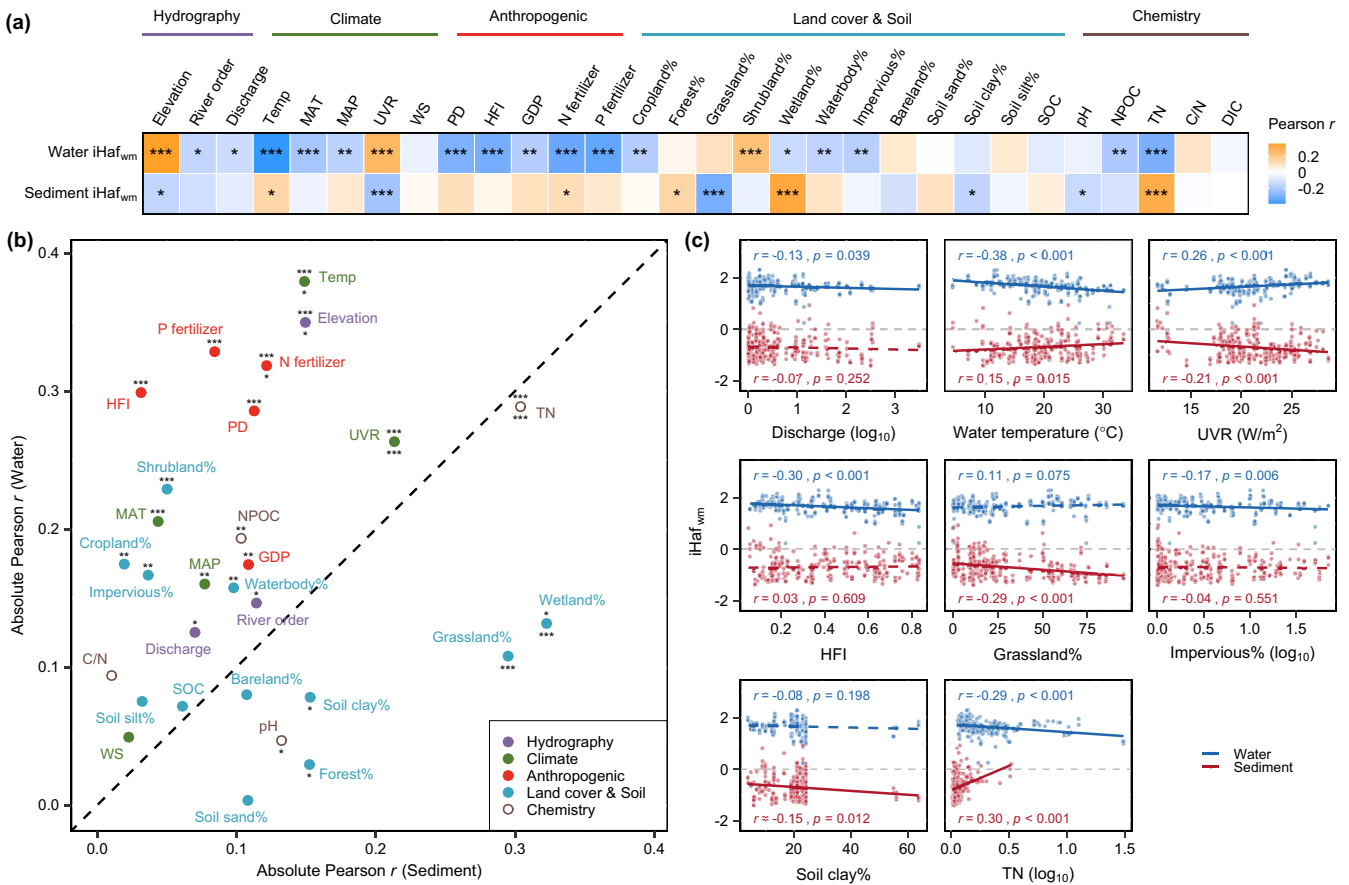


FIGURE 4 | Potential environmental drivers of compositional-level habitat affinity ($iHaf_{wm}$) of DOM. (a) Pearson correlation coefficients of the habitat affinities for DOM assemblages with environmental variables. (b) Comparison of absolute Pearson correlations between DOM habitat affinity and environmental variables in waters and sediments. Environmental variables are grouped by hydrography, climate, anthropogenic influences, land cover and soil, and chemistry variables with different colors. Dotted lines indicate the 1:1. Asterisks above and below the circles indicate significant ($p \leq 0.05$) relationships between $iHaf_{wm}$ and environmental variables in waters and sediments, respectively. The asterisks in (a) and (b) indicate the significance levels as follows: *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. (c) Variations in habitat affinities of DOM assemblages with environmental variables. Solid and dashed lines indicate statistically significant ($p \leq 0.05$) and non-significant ($p > 0.05$) relationships, respectively. Abbreviations of environmental variables are detailed in Table S1.

waters and a 1.5% increase in sediments, indicating that higher temperature is associated with weaker habitat affinity and could enhance molecular homogenization between these two habitats. In contrast, the $iHaf_{wm}$ in waters was positively correlated with ultraviolet radiation (UVR; $r = 0.26$, $p < 0.001$), while the $iHaf_{wm}$ in sediments was negatively correlated ($r = -0.21$, $p < 0.001$) (Figure 4), indicating that photodegradation could increase DOM habitat affinity. Sunlight promotes the degradation of condensed aromatic-like compounds (An et al. 2024; Milstead et al. 2023), which were identified as low-habitat-affinity molecules (Figure 1d,e), thereby reducing the likelihood of molecular exchange between waters and sediments.

Elevated temperatures likely increase the kinetic energy of DOM molecules (Davidson and Janssens 2006; Hu et al. 2024), thereby promoting desorption, diffusion, and resuspension processes that enhance molecular redistribution across habitats (Cornelissen et al. 1997; Dadi et al. 2017). Experimental evidence has demonstrated that the mobilization rate coefficient of aqueous-phase organic molecules increases with increasing temperature (Xu and Saiers 2010). As global warming enhances molecular connectivity between habitats, it may stimulate

microbial substrate availability and ultimately accelerate DOM turnover and carbon processing in riverine systems (Battin et al. 2023; Hu, Choi, et al. 2022). Consequently, the joint effects of enhanced physical transport and accelerated microbial turnover decrease the compositional divergence between water and sediment habitats, leading to molecular homogenization.

Anthropogenic disturbances and land-use changes in the surrounding landscape also exert a strong influence on DOM habitat affinity. $iHaf_{wm}$ in waters was negatively correlated with human footprint index (HFI; $r = -0.30$, $p < 0.001$), N fertilizer use rate ($r = -0.32$, $p < 0.001$), impervious surface area ($r = -0.17$, $p = 0.006$), cropland coverage ($r = -0.17$, $p = 0.004$), and water total nitrogen ($r = -0.29$, $p < 0.001$), but positively correlated with shrubland ($r = 0.23$, $p < 0.001$) (Figure 4). In contrast, $iHaf_{wm}$ in sediments was positively correlated with total nitrogen concentration ($r = 0.30$, $p < 0.001$), but negatively correlated with grassland coverage ($r = -0.29$, $p < 0.001$) and soil clay content ($r = -0.15$, $p = 0.012$) (Figure 4). These results indicate that natural land covers such as shrubland and grassland promote the selective accumulation of DOM in distinct habitats. However, land-use intensification, particularly conversion to

cropland or urban impervious surfaces, reduces habitat affinity for both waters and sediments and could enhance molecular exchange across habitats.

Anthropogenic pressures such as nitrogen loading likely stimulate microbial and algal productivity (Hu et al. 2024; Wang et al. 2016), increasing the input of labile carbon to sediments and enhancing molecular flux between habitats. Furthermore, impervious surfaces and coarse-textured soils (e.g., low clay content) promote rapid runoff and sediment transport, intensifying hydrological connectivity and enhancing DOM transport across the sediment–water interface (Awad et al. 2015; Harjung et al. 2018; Jacobson 2011). This is further supported by the hydrographic factors such as river order ($r = -0.15$, $p = 0.016$) and river discharge ($r = -0.13$, $p = 0.039$), which were negatively correlated with $iHaf_{wm}$ in waters (Figure 4), indicating that large stream size and high discharge weaken DOM habitat affinity. Episodic high-flow events in urban catchments further enhance physical disturbance, resuspending benthic materials and mobilizing sediment-bound DOM into the overlying water column (Lu and Liu 2019; McCabe et al. 2021; Raymond et al. 2016). Collectively, our results suggest that both warming and anthropogenic pressures increase DOM exchange potential and possibly accelerate riverine carbon loss, as indicated by the positive correlation between the relative abundance of low-habitat-affinity molecules and sediment respiration rates (Figure S5). Instead, watershed restoration through the reduction of cropland expansion and the reestablishment of natural vegetation could enhance the DOM persistence and support long-term riverine carbon conservation.

3.5 | Global Mapping of Habitat Affinities for Riverine DOM Assemblages

We applied a machine learning model to provide an exploratory prediction for the habitat affinities of DOM assemblages across global rivers, using 43 environmental variables, including hydrography, climate, anthropogenic influences, land cover, and soil properties (Table S1; see Section 2). The random forest model showed strong predictive performance, explaining 69% and 61% of the variation for waters and sediments, respectively (Figure S9). For waters, N fertilizer use rate was the most influential predictor, and other anthropogenic pressures such as population density, cropland coverage, and impervious surface area were also important (Figure S9). For sediments, wetland coverage was most important, followed by elevation, mean temperature of wettest quarter, and precipitation of warmest quarter (Figure S9).

The projected global patterns of DOM habitat affinity revealed distinct spatial heterogeneity, with low absolute affinity values—indicating high molecular exchange potential—occurring predominantly in regions characterized by intense anthropogenic disturbances and warm climates. On average, global $iHaf_{wm}$ values were $1.62 (\pm 0.07; SD)$ for waters and $-0.58 (\pm 0.16)$ for sediments (Figure 5). In waters, $iHaf_{wm}$ varied markedly with latitude, with lowest values (i.e., highest cross-habitat exchange potential) at around $24^\circ N$, while in sediments, the lowest absolute values were observed near the equator ($\sim 2^\circ S$) and in

high-latitude regions of the Northern Hemisphere (Figure 5b,d). Hotspots of low habitat affinity for waters were mainly distributed in regions with concentrated nitrogen fertilizer use and intense human activities, including East and Southeast Asia and Western Europe (Figure 5a and Figure S10). In contrast, hotspots of low habitat affinity for sediments were mostly in tropical regions, such as Southeast Asia and the Amazon basin (Figure 5c and Figure S10).

The above contrasting spatial patterns of DOM habitat affinities between river waters and sediments may be caused by the differential influence of anthropogenic activities across climate zones. As expected, the cross-habitat exchange of DOM should increase with increasing temperature for both waters and sediments (Figure 4). However, our global patterns showed that in waters, the highest exchange potential could occur in temperate rather than tropical zones (Figure 5a). In contrast, the pattern of DOM habitat affinities in sediments aligns with the climatic expectations, showing the highest molecular exchange potential in tropical zones (Figure 5c). Such an observation is consistent with previous findings of the weaker anthropogenic influences on sedimentary environments compared with overlying waters (Cui et al. 2024), or the promoted deoxygenation or hypoxia in warmer rivers which facilitates the release of DOM molecules from sediments into the water column (Peter et al. 2016; Riedel et al. 2013; Zhi et al. 2023). Collectively, these results indicate that while warming enhances molecular exchange potential across riverine habitats, the extent of this response diverges between waters and sediments across climate zones, emphasizing the important influence of anthropogenic pressures on the biogeochemical functioning of fluvial systems.

We encourage future research to apply the current novel framework by including temporal information, biotic drivers, and chemical structures of DOM. For instance, when field logistics are allowed, multi-seasonal or high-frequency sampling at global scales and across extreme hydrological events such as flash floods could be conducted to capture the potential temporal variability in DOM habitat affinity. Well-designed laboratory or in situ experiments, by integrating microbial community compositions and ecosystem functions (Hu et al. 2024; Wu et al. 2024), would further help support the mechanisms driving molecular exchange potential at the sediment–water interface, to disentangle the underlying ecological processes, and to link the DOM habitat affinity to ecosystem functioning. It should be noted that given the limitations of FT-ICR MS in deriving chemical structures, other analytical techniques, such as liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy (Cai et al. 2025; Zark and Dittmar 2018), are useful to identify the isomeric composition of DOM and their associated biogeochemical processes in future studies.

In summary, our study reveals how, and to what extent, complex and diverse DOM molecules exhibit habitat affinities, that is, are preferentially accumulated in either waters or sediments, within riverine ecosystems. By integrating large-scale spatial observations with a newly developed quantitative approach, we show that most molecules exhibit low habitat affinity, indicative of higher exchange potential, and tend to be

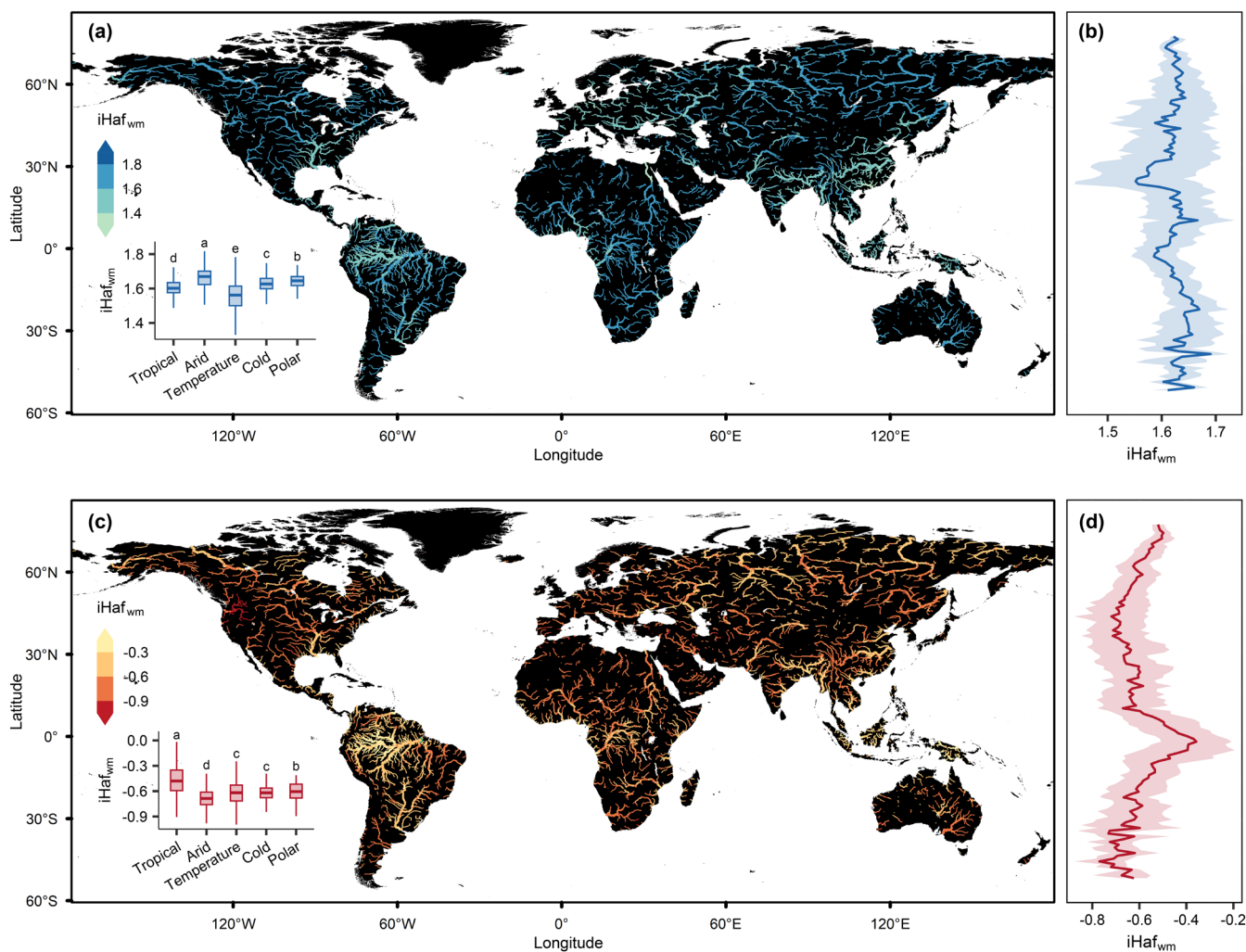


FIGURE 5 | Habitat affinities of DOM assemblages ($iHaf_{wm}$) mapped across the global rivers. (a, c) The spatial distribution of estimated DOM habitat affinities for (a) waters and (c) sediments in the global river network. Insets show the distribution of $iHaf_{wm}$ across different climate zones. Boxes indicate interquartile range, and thick horizontal lines within boxes denote median values. Different letters indicate significant ($p \leq 0.05$) differences based on the Kruskal-Wallis test. (b, d) Latitudinal patterns of the habitat affinities of riverine DOM assemblages for (b) waters and (d) sediments. The solid line denotes the latitudinal mean values within 1° latitude bins, and shading denotes their standard deviations. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

more thermodynamically favorable and more susceptible to microbial degradation. Our indicator of habitat affinity ($iHaf$) can be upscaled to the entire DOM assemblage, and the values approaching zero indicate homogenized molecular composition between these two habitats partly due to high exchange potential. Importantly, our findings suggest that future warming and intensified anthropogenic activities, such as population growth, nutrient enrichment, and land-use change, reduce the habitat affinity of DOM in both waters and sediments. This shift enhances molecular exchange potential between habitats and may increase the vulnerability of riverine carbon pools to loss. These insights underscore the need to integrate DOM habitat affinity into predictive carbon-climate models and mitigation strategies. Accounting for this molecular-level mechanism is critical for improving predictions of carbon cycling under climate warming and for informing more effective carbon conservation policies. Notably, reducing cropland expansion and restoring natural landscapes within watersheds may enhance the persistence of DOM and improve global riverine carbon conservation. Looking

ahead, we propose that the habitat affinity indicator for DOM assemblages can be broadly applied across diverse ecosystems to assess and predict DOM turnover along geographical and environmental gradients.

Author Contributions

J.W. conceived the idea. Y.C., A.H., and J.W. performed the statistical analyses. Y.C. wrote the first draft of the manuscript. A.H. and J.W. finished the manuscript with comments from J.C.S. All authors contributed to the intellectual development of this study.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (42225708, U24A20578, 42377122, 42401135), Basic Research Program of Jiangsu (BK20240111), Key Laboratory of Lake and Watershed Science for Water Security (NKL2023-QN04), and Science and Technology Planning Project of NIGLAS (NIGLAS2022GS09). We appreciate the WHONDRS consortium, including the team at Pacific

Northwest National Laboratory, for field sampling, sample processing, and data publication. We appreciate Shaoda Liu, Guoyao Huang, and Shuailong Wen for kind support.

Funding

This work was supported by the National Natural Science Foundation of China, 42225708, U24A20578, 42377122, and 42401135. Basic Research Program of Jiangsu Province, BK20240111. Key Laboratory of Lake and Watershed Science for Water Security, NKL2023-QN04. Science and Technology Planning Project of NIGLAS, NIGLAS2022GS09.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

DOM and related environmental data for river water and sediment samples are available at the ESS-DIVE (<https://doi.org/10.15485/1603775>; <https://doi.org/10.15485/1729719>).

References

An, S., Y. Du, X. Huang, et al. 2024. “Long-Term Photochemical and Microbial Alterations Lead to the Compositional Convergence of Algal and Terrestrial Dissolved Organic Matter.” *Environmental Science & Technology* 58, no. 42: 18765–18776. <https://doi.org/10.1021/acs.est.4c07307>.

Arianoutsou, M., I. Bazos, P. Delipetrou, and Y. Kokkoris. 2010. “The Alien Flora of Greece: Taxonomy, Life Traits and Habitat Preferences.” *Biological Invasions* 12, no. 10: 3525–3549. <https://doi.org/10.1007/s10530-010-9749-0>.

Aukes, P. J. K., and S. L. Schiff. 2021. “Composition Wheels: Visualizing Dissolved Organic Matter Using Common Composition Metrics Across a Variety of Canadian Ecozones.” *PLoS One* 16, no. 7: e0253972. <https://doi.org/10.1371/journal.pone.0253972>.

Awad, J., J. van Leeuwen, D. Abate, et al. 2015. “The Effect of Vegetation and Soil Texture on the Nature of Organics in Runoff From a Catchment Supplying Water for Domestic Consumption.” *Science of the Total Environment* 529: 72–81. <https://doi.org/10.1016/j.scitotenv.2015.05.037>.

Bastian, M., S. Heymann, and M. Jacomy. 2009. “Gephi: An Open Source Software for Exploring and Manipulating Networks.” *Proceedings of the International AAAI Conference on Web and Social Media* 3, no. 1: 361–362. <https://doi.org/10.1609/icwsm.v3i1.13937>.

Battin, T. J., L. A. Kaplan, S. Findlay, et al. 2008. “Biophysical Controls on Organic Carbon Fluxes in Fluvial Networks.” *Nature Geoscience* 1, no. 2: 95–100. <https://doi.org/10.1038/ngeo101>.

Battin, T. J., R. Lauerwald, E. S. Bernhardt, et al. 2023. “River Ecosystem Metabolism and Carbon Biogeochemistry in a Changing World.” *Nature* 613, no. 7944: 449–459. <https://doi.org/10.1038/s41586-022-05500-8>.

Battin, T. J., S. Luysaert, L. A. Kaplan, A. K. Aufdenkampe, A. Richter, and L. J. Tranvik. 2009. “The Boundless Carbon Cycle.” *Nature Geoscience* 2, no. 9: 598–600. <https://doi.org/10.1038/ngeo618>.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood. 2018. “Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution.” *Scientific Data* 5, no. 1: 180214. <https://doi.org/10.1038/sdata.2018.214>.

Beyer, H. L., D. T. Haydon, J. M. Morales, et al. 2010. “The Interpretation of Habitat Preference Metrics Under Use–Availability Designs.” *Philosophical Transactions of the Royal Society, B: Biological Sciences* 365, no. 1550: 2245–2254. <https://doi.org/10.1098/rstb.2010.0083>.

Cai, R., O. J. Lechtenfeld, Z. Yan, et al. 2025. “Constraining Biorecalcitrance of Carboxyl-Rich Alicyclic Molecules in the Ocean.”

Science Advances 11, no. 28: eadw1148. <https://doi.org/10.1126/sciadv.adw1148>.

Catalán, N., R. Marcé, D. N. Kothawala, and L. J. Tranvik. 2016. “Organic Carbon Decomposition Rates Controlled by Water Retention Time Across Inland Waters.” *Nature Geoscience* 9, no. 7: 501–504. <https://doi.org/10.1038/ngeo2720>.

Chen, Q., J. Chen, R. Cai, et al. 2025. “Deciphering Microbially Driven Labile and Refractory Molecular Candidates in Dissolved Organic Matter.” *Journal of Geophysical Research: Biogeosciences* 130, no. 2: e2024JG008484. <https://doi.org/10.1029/2024JG008484>.

Cornelissen, G., P. C. M. van Noort, J. R. Parsons, and H. A. J. Govers. 1997. “Temperature Dependence of Slow Adsorption and Desorption Kinetics of Organic Compounds in Sediments.” *Environmental Science & Technology* 31, no. 2: 454–460. <https://doi.org/10.1021/es960300>.

Creed, I. F., D. M. McKnight, B. A. Pellerin, et al. 2015. “The River as a Chemostat: Fresh Perspectives on Dissolved Organic Matter Flowing Down the River Continuum.” *Canadian Journal of Fisheries and Aquatic Sciences* 72, no. 8: 1272–1285. <https://doi.org/10.1139/cjfas-2014-0400>.

Csárdi, G., and T. Nepusz. 2006. “The Igraph Software Package for Complex Network Research.” *Interjournal Complex Systems* 1695, no. 5: 1–9.

Cui, Y., S. Wen, J. C. Stegen, A. Hu, and J. Wang. 2024. “Chemodiversity of Riverine Dissolved Organic Matter: Effects of Local Environments and Watershed Characteristics.” *Water Research* 250: 121054. <https://doi.org/10.1016/j.watres.2023.121054>.

Cui, Y., S. Wen, J. C. Stegen, A. Hu, and J. Wang. 2026. “Unveiling the Dark Matter of Riverine Dissolved Organic Matter and Its Role in Molecular Chemodiversity.” *Water Research* 289: 124870. <https://doi.org/10.1016/j.watres.2025.124870>.

Dadi, T., K. Wendt-Potthoff, and M. Koschorreck. 2017. “Sediment Resuspension Effects on Dissolved Organic Carbon Fluxes and Microbial Metabolic Potentials in Reservoirs.” *Aquatic Sciences* 79, no. 3: 749–764. <https://doi.org/10.1007/s00027-017-0533-4>.

Danczak, R. E., A. E. Goldman, R. K. Chu, et al. 2021. “Ecological Theory Applied to Environmental Metabolomes Reveals Compositional Divergence Despite Conserved Molecular Properties.” *Science of the Total Environment* 788: 147409. <https://doi.org/10.1016/j.scitotenv.2021.147409>.

D’Andrilli, J., W. T. Cooper, C. M. Foreman, and A. G. Marshall. 2015. “An Ultrahigh-Resolution Mass Spectrometry Index to Estimate Natural Organic Matter Lability.” *Rapid Communications in Mass Spectrometry* 29, no. 24: 2385–2401. <https://doi.org/10.1002/rcm.7400>.

Davidson, E. A., and I. A. Janssens. 2006. “Temperature Sensitivity of Soil Carbon Decomposition and Feedbacks to Climate Change.” *Nature* 440, no. 7081: 165–173. <https://doi.org/10.1038/nature04514>.

Delgado-Baquerizo, M., A. M. Oliverio, T. E. Brewer, et al. 2018. “A Global Atlas of the Dominant Bacteria Found in Soil.” *Science* 359, no. 6373: 320–325. <https://doi.org/10.1126/science.aap9516>.

Dittmar, T., S. T. Lennartz, H. Buck-Wiese, et al. 2021. “Enigmatic Persistence of Dissolved Organic Matter in the Ocean.” *Nature Reviews Earth & Environment* 2, no. 8: 570–583. <https://doi.org/10.1038/s43017-021-00183-7>.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. “A Working Guide to Boosted Regression Trees.” *Journal of Animal Ecology* 77, no. 4: 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.

Felipe-Lucia, M. R., S. Soliveres, C. Penone, et al. 2020. “Land-Use Intensity Alters Networks Between Biodiversity, Ecosystem Functions, and Services.” *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 45: 28140–28149. <https://doi.org/10.1073/pnas.2016210117>.

- Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37, no. 12: 4302–4315. <https://doi.org/10.1002/joc.5086>.
- Fischer, G., F. Nachtergaele, S. Prieler, H. Van Velthuizen, L. Verelst, and D. Wiberg. 2008. *Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, FAO.
- Freeman, E. C., E. J. S. Emilson, T. Dittmar, et al. 2024. "Universal Microbial Reworking of Dissolved Organic Matter Along Environmental Gradients." *Nature Communications* 15, no. 1: 187. <https://doi.org/10.1038/s41467-023-44431-4>.
- Garayburu-Caruso, V. A., R. E. Danczak, J. C. Stegen, et al. 2020. "Using Community Science to Reveal the Global Chemogeography of River Metabolomes." *Metabolites* 10, no. 12: 518. <https://doi.org/10.3390/metabo10120518>.
- Goldman, A. E., S. Arnon, E. Bar-Zeev, et al. 2020. *WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Sediment FTICR-MS, Dissolved Organic Carbon, Aerobic Respiration, Elemental Composition, Grain Size, Total Nitrogen and Organic Carbon Content, Bacterial Abundance, and Stable Isotopes*. U.S. Department of Energy, Office of Scientific and Technical Information. <https://doi.org/10.15485/1729719>.
- Groeneveld, M., N. Catalán, K. Attermeyer, et al. 2020. "Selective Adsorption of Terrestrial Dissolved Organic Matter to Inorganic Surfaces Along a Boreal Inland Water Continuum." *Journal of Geophysical Research: Biogeosciences* 125, no. 3: e2019JG005236. <https://doi.org/10.1029/2019JG005236>.
- Hansell, D. A., and C. A. Carlson. 2014. *Biogeochemistry of Marine Dissolved Organic Matter*. Academic Press.
- Harjung, A., F. Sabater, and A. Butturini. 2018. "Hydrological Connectivity Drives Dissolved Organic Matter Processing in an Intermittent Stream." *Limnologia* 68: 71–81. <https://doi.org/10.1016/j.limno.2017.02.007>.
- Harvey, J., and M. Gooseff. 2015. "River Corridor Science: Hydrologic Exchange and Ecological Consequences From Bedforms to Basins." *Water Resources Research* 51, no. 9: 6893–6922. <https://doi.org/10.1002/2015WR017617>.
- Hedges, L. V. 1981. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics* 6, no. 2: 107–128. <https://doi.org/10.3102/10769986006002107>.
- Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. Academic Press.
- Hersbach, H., B. Bell, P. Berrisford, et al. 2023. *ERA5 Monthly Averaged Data on Single Levels From 1940 to Present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.f17050d7>.
- Hu, A., M. Choi, A. J. Tanentzap, et al. 2022. "Ecological Networks of Dissolved Organic Matter and Microorganisms Under Global Change." *Nature Communications* 13, no. 1: 3600. <https://doi.org/10.1038/s41467-022-31251-1>.
- Hu, A., K.-S. Jang, F. Meng, et al. 2022. "Microbial and Environmental Processes Shape the Link Between Organic Matter Functional Traits and Composition." *Environmental Science & Technology* 56, no. 14: 10504–10516. <https://doi.org/10.1021/acs.est.2c01432>.
- Hu, A., K.-S. Jang, A. J. Tanentzap, et al. 2024. "Thermal Responses of Dissolved Organic Matter Under Global Change." *Nature Communications* 15, no. 1: 576. <https://doi.org/10.1038/s41467-024-44813-2>.
- Hu, A., J. Stegen, A. J. Tanentzap, and J. Wang. 2025. "The Emergence and Promise of Functional Chemogeography of Organic Matter." *Global Change Biology* 31, no. 8: e70435. <https://doi.org/10.1111/gcb.70435>.
- Jacobson, C. R. 2011. "Identification and Quantification of the Hydrological Impacts of Imperviousness in Urban Catchments: A Review." *Journal of Environmental Management* 92, no. 6: 1438–1448. <https://doi.org/10.1016/j.jenvman.2011.01.018>.
- Johnson, D. H. 1980. "The Comparison of Usage and Availability Measurements for Evaluating Resource Preference." *Ecology* 61, no. 1: 65–71. <https://doi.org/10.2307/1937156>.
- Kellerman, A. M., T. Dittmar, D. N. Kothawala, and L. J. Tranvik. 2014. "Chemodiversity of Dissolved Organic Matter in Lakes Driven by Climate and Hydrology." *Nature Communications* 5, no. 1: 3804. <https://doi.org/10.1038/ncomms4804>.
- Kew, W., A. Myers-Pigg, C. H. Chang, et al. 2024. "Reviews and Syntheses: Opportunities for Robust Use of Peak Intensities From High-Resolution Mass Spectrometry in Organic Matter Studies." *Biogeosciences* 21, no. 20: 4665–4679. <https://doi.org/10.5194/bg-21-4665-2024>.
- Keys, P. W., E. A. Barnes, and N. H. Carter. 2021. "A Machine-Learning Approach to Human Footprint Index Estimation With Applications to Sustainable Development." *Environmental Research Letters* 16, no. 4: 44061. <https://doi.org/10.1088/1748-9326/abe00a>.
- Kido Soule, M. C., K. Longnecker, S. J. Giovannoni, and E. B. Kujawinski. 2010. "Impact of Instrument and Experiment Parameters on Reproducibility of Ultrahigh Resolution ESI FT-ICR Mass Spectra of Natural Organic Matter." *Organic Geochemistry* 41, no. 8: 725–733. <https://doi.org/10.1016/j.orggeochem.2010.05.017>.
- Kim, S., R. W. Kramer, and P. G. Hatcher. 2003. "Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram." *Analytical Chemistry* 75, no. 20: 5336–5344. <https://doi.org/10.1021/ac034415p>.
- Koch, B. P., and T. Dittmar. 2006. "From Mass to Structure: An Aromaticity Index for High-Resolution Mass Data of Natural Organic Matter." *Rapid Communications in Mass Spectrometry* 20, no. 5: 926–932. <https://doi.org/10.1002/rcm.2386>.
- Kong, X., O. J. Lechtenfeld, J. M. Kaesler, et al. 2025. "Major Terrestrial Contribution to the Dissolved Organic Carbon Budget in the Arctic Ocean." *Nature Geoscience* 19, no. 1: 90–98. <https://doi.org/10.1038/s41561-025-01847-5>.
- Kummu, M., M. Taka, and J. H. A. Guillaume. 2018. "Gridded Global Datasets for Gross Domestic Product and Human Development Index Over 1990–2015." *Scientific Data* 5, no. 1: 180004. <https://doi.org/10.1038/sdata.2018.4>.
- LaRowe, D. E., and P. Van Cappellen. 2011. "Degradation of Natural Organic Matter: A Thermodynamic Analysis." *Geochimica et Cosmochimica Acta* 75, no. 8: 2030–2042. <https://doi.org/10.1016/j.gca.2011.01.020>.
- Lavorel, S., K. Grigulis, S. McIntyre, et al. 2008. "Assessing Functional Diversity in the Field—Methodology Matters!" *Functional Ecology* 22, no. 1: 134–147. <https://doi.org/10.1111/j.1365-2435.2007.01339.x>.
- Lehmann, J., and M. Kleber. 2015. "The Contentious Nature of Soil Organic Matter." *Nature* 528, no. 7580: 60–68. <https://doi.org/10.1038/nature16069>.
- Lehner, B., and G. Grill. 2013. "Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems." *Hydrological Processes* 27, no. 15: 2171–2186. <https://doi.org/10.1002/hyp.9740>.
- Li, A., X. Tian, N. Ma, and J. Luo. 2025. "Habitat Preference Contributes to Explaining the Varied Sensitivity of Bats to Anthropogenic Noise." *Biological Conservation* 302: 110974. <https://doi.org/10.1016/j.biocon.2025.110974>.
- Li, C., P. Gong, J. Wang, et al. 2017. "The First All-Season Sample Set for Mapping Global Land Cover With Landsat-8 Data." *Science Bulletin* 62, no. 7: 508–515. <https://doi.org/10.1016/j.scib.2017.03.011>.

- Lin, P., M. Pan, H. E. Beck, et al. 2019. "Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches." *Water Resources Research* 55, no. 8: 6499–6516. <https://doi.org/10.1029/2019WR025287>.
- Lu, C., and H. Tian. 2017. "Global Nitrogen and Phosphorus Fertilizer Use for Agriculture Production in the Past Half Century: Shifted Hot Spots and Nutrient Imbalance." *Earth System Science Data* 9, no. 1: 181–192. <https://doi.org/10.5194/essd-9-181-2017>.
- Lu, K., and Z. Liu. 2019. "Molecular Level Analysis Reveals Changes in Chemical Composition of Dissolved Organic Matter From South Texas Rivers After High Flow Events." *Frontiers in Marine Science* 6: 673. <https://doi.org/10.3389/fmars.2019.00673>.
- Lynch, L. M., N. A. Sutfin, T. S. Feghel, C. M. Boot, T. P. Covino, and M. D. Wallenstein. 2019. "River Channel Connectivity Shifts Metabolite Composition and Dissolved Organic Matter Chemistry." *Nature Communications* 10, no. 1: 459. <https://doi.org/10.1038/s41467-019-08406-8>.
- Maire, V., N. Gross, L. Börger, et al. 2012. "Habitat Filtering and Niche Differentiation Jointly Explain Species Relative Abundance Within Grassland Communities Along Fertility and Disturbance Gradients." *New Phytologist* 196, no. 2: 497–509. <https://doi.org/10.1111/j.1469-8137.2012.04287.x>.
- McCabe, K. M., E. M. Smith, S. Q. Lang, C. L. Osburn, and C. R. Benitez-Nelson. 2021. "Particulate and Dissolved Organic Matter in Stormwater Runoff Influences Oxygen Demand in Urbanized Headwater Catchments." *Environmental Science & Technology* 55, no. 2: 952–961. <https://doi.org/10.1021/acs.est.0c04502>.
- McDonough, L. K., M. S. Andersen, M. I. Behnke, et al. 2022. "A New Conceptual Framework for the Transformation of Groundwater Dissolved Organic Matter." *Nature Communications* 13, no. 1: 2153. <https://doi.org/10.1038/s41467-022-29711-9>.
- Meng, F., A. Hu, K.-S. Jang, and J. Wang. 2025. "iDOM: Statistical Analysis of Dissolved Organic Matter Characterized by High-Resolution Mass Spectrometry." *mLife* 4: 319–331. <https://doi.org/10.1002/mlf2.70002>.
- Milstead, R. P., E. R. Horvath, and C. K. Remucal. 2023. "Dissolved Organic Matter Composition Determines Its Susceptibility to Complete and Partial Photooxidation Within Lakes." *Environmental Science & Technology* 57, no. 32: 11876–11885. <https://doi.org/10.1021/acs.est.3c01500>.
- Naimi, B., and M. B. Araújo. 2016. "Sdm: A Reproducible and Extensible R Platform for Species Distribution Modelling." *Ecography* 39, no. 4: 368–375. <https://doi.org/10.1111/ecog.01881>.
- Peter, S., A. Isidorova, and S. Sobek. 2016. "Enhanced Carbon Loss From Anoxic Lake Sediment Through Diffusion of Dissolved Organic Carbon." *Journal of Geophysical Research: Biogeosciences* 121, no. 7: 1959–1977. <https://doi.org/10.1002/2016JG003425>.
- Picazo, F., A. Vilmi, J. Aalto, et al. 2020. "Climate Mediates Continental Scale Patterns of Stream Microbial Functional Diversity." *Microbiome* 8, no. 1: 92. <https://doi.org/10.1186/s40168-020-00873-2>.
- Raymond, P. A., J. E. Saiers, and W. V. Sobczak. 2016. "Hydrological and Biogeochemical Controls on Watershed Dissolved Organic Matter Transport: Pulse-Shunt Concept." *Ecology* 97, no. 1: 5–16. <https://doi.org/10.1890/14-1684.1>.
- Regnier, P., P. Friedlingstein, P. Ciais, et al. 2013. "Anthropogenic Perturbation of the Carbon Fluxes From Land to Ocean." *Nature Geoscience* 6, no. 8: 597–607. <https://doi.org/10.1038/ngeo1830>.
- Regnier, P., L. Resplandy, R. G. Najjar, and P. Ciais. 2022. "The Land-To-Ocean Loops of the Global Carbon Cycle." *Nature* 603, no. 7901: 401–410. <https://doi.org/10.1038/s41586-021-04339-9>.
- Riedel, T., D. Zak, H. Biester, and T. Dittmar. 2013. "Iron Traps Terrestrially Derived Dissolved Organic Matter at Redox Interfaces." *Proceedings of the National Academy of Sciences* 110, no. 25: 10101–10105. <https://doi.org/10.1073/pnas.1221487110>.
- Riedel, T., M. Zark, A. V. Vähätalo, et al. 2016. "Molecular Signatures of Biogeochemical Transformations in Dissolved Organic Matter From Ten World Rivers." *Frontiers in Earth Science* 4: 85. <https://doi.org/10.3389/feart.2016.00085>.
- Roth, V.-N., M. Lange, C. Simon, et al. 2019. "Persistence of Dissolved Organic Matter Explained by Molecular Changes During Its Passage Through Soil." *Nature Geoscience* 12, no. 9: 755–761. <https://doi.org/10.1038/s41561-019-0417-4>.
- Šantl-Temkiv, T., K. Finster, T. Dittmar, et al. 2013. "Hailstones: A Window Into the Microbial and Chemical Inventory of a Storm Cloud." *PLoS One* 8, no. 1: e53550. <https://doi.org/10.1371/journal.pone.0053550>.
- Santoemma, G., F. Trivellato, V. Caloi, N. Mori, and L. Marini. 2019. "Habitat Preference of *Drosophila Suzukii* Across Heterogeneous Landscapes." *Journal of Pest Science* 92, no. 2: 485–494. <https://doi.org/10.1007/s10340-018-1052-3>.
- Scherer-Lorenzen, M., M. O. Gessner, B. E. Beisner, et al. 2022. "Pathways for Cross-Boundary Effects of Biodiversity on Ecosystem Functioning." *Trends in Ecology & Evolution* 37, no. 5: 454–467. <https://doi.org/10.1016/j.tree.2021.12.009>.
- Schindler, D. E., and M. D. Scheuerell. 2002. "Habitat Coupling in Lake Ecosystems." *Oikos* 98, no. 2: 177–189. <https://doi.org/10.1034/j.1600-0706.2002.980201.x>.
- Song, H.-S., J. C. Stegen, E. B. Graham, et al. 2020. "Representing Organic Matter Thermodynamics in Biogeochemical Reactions via Substrate-Explicit Modeling." *Frontiers in Microbiology* 11: 531756. <https://doi.org/10.3389/fmicb.2020.531756>.
- Sowers, T. D., K. L. Holden, E. K. Coward, and D. L. Sparks. 2019. "Dissolved Organic Matter Sorption and Molecular Fractionation by Naturally Occurring Bacteriogenic Iron (Oxyhydr)oxides." *Environmental Science & Technology* 53, no. 8: 4295–4304. <https://doi.org/10.1021/acs.est.9b00540>.
- Stadler, M., M. A. Barnard, K. Bice, et al. 2023. "Applying the Core-Satellite Species Concept: Characteristics of Rare and Common Riverine Dissolved Organic Matter." *Frontiers in Water* 5: 1156042. <https://doi.org/10.3389/frwa.2023.1156042>.
- Stegen, J. C., S. J. Fansler, M. M. Tfaily, et al. 2022. "Organic Matter Transformations Are Disconnected Between Surface Water and the Hyporheic Zone." *Biogeosciences* 19, no. 12: 3099–3110. <https://doi.org/10.5194/bg-19-3099-2022>.
- Stegen, J. C., and A. E. Goldman. 2018. "WHONDORS: A Community Resource for Studying Dynamic River Corridors." *mSystems* 3, no. 5: e00151-00118. <https://doi.org/10.1128/msystems.00151-18>.
- Stegen, J. C., T. Johnson, J. K. Fredrickson, et al. 2018. "Influences of Organic Carbon Speciation on Hyporheic Corridor Biogeochemistry and Microbial Ecology." *Nature Communications* 9, no. 1: 585. <https://doi.org/10.1038/s41467-018-02922-9>.
- Toyoda, J. G., A. E. Goldman, S. Arnon, et al. 2020. "WHONDORS Summer 2019 Sampling Campaign: Global River Corridor Surface Water FTICR-MS, NPOC, TN, Anions, Stable Isotopes, Bacterial Abundance, and Dissolved Inorganic Carbon." <https://doi.org/10.15485/1603775>.
- Tranvik, L. J., J. A. Downing, J. B. Cotner, et al. 2009. "Lakes and Reservoirs as Regulators of Carbon Cycling and Climate." *Limnology and Oceanography* 54: 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298.
- Viechtbauer, W. 2010. "Conducting Meta-Analyses in R With the Metafor Package." *Journal of Statistical Software* 36, no. 3: 1–48. <https://doi.org/10.18637/jss.v036.i03>.

Wang, J., F. Pan, J. Soininen, J. Heino, and J. Shen. 2016. "Nutrient Enrichment Modifies Temperature-Biodiversity Relationships in Large-Scale Field Experiments." *Nature Communications* 7: 13960. <https://doi.org/10.1038/ncomms13960>.

Wang, J., J. Shen, Y. Wu, et al. 2013. "Phylogenetic Beta Diversity in Bacterial Assemblages Across Ecosystems: Deterministic Versus Stochastic Processes." *ISME Journal* 7, no. 7: 1310–1321. <https://doi.org/10.1038/ismej.2013.30>.

Wang, M., S. Zhang, X. Guo, et al. 2024. "Responses of Soil Organic Carbon to Climate Extremes Under Warming Across Global Biomes." *Nature Climate Change* 14, no. 1: 98–105. <https://doi.org/10.1038/s41558-023-01874-3>.

Wilson, H. F., and M. A. Xenopoulos. 2008. "Effects of Agricultural Land Use on the Composition of Fluvial Dissolved Organic Matter." *Nature Geoscience* 2, no. 1: 37–41. <https://doi.org/10.1038/ngeo391>.

Wu, H., T. Gao, A. Hu, and J. Wang. 2024. "Network Complexity and Stability of Microbes Enhanced by Microplastic Diversity." *Environmental Science & Technology* 58, no. 9: 4334–4345. <https://doi.org/10.1021/acs.est.3c08704>.

Xu, N., and J. E. Saiers. 2010. "Temperature and Hydrologic Controls on Dissolved Organic Matter Mobilization and Transport Within a Forest Topsoil." *Environmental Science & Technology* 44, no. 14: 5423–5429. <https://doi.org/10.1021/es1002296>.

Yamazaki, D., D. Ikeshima, R. Tawatari, et al. 2017. "A High-Accuracy Map of Global Terrain Elevations." *Geophysical Research Letters* 44, no. 11: 5844–5853. <https://doi.org/10.1002/2017GL072874>.

Yi, Y., A. J. Tanentzap, C. He, et al. 2025. "Underestimated Input of Terrestrial Dissolved Organic Carbon to the Ocean." *Proceedings of the National Academy of Sciences of the United States of America* 122, no. 45: e2505838122. <https://doi.org/10.1073/pnas.2505838122>.

Zark, M., and T. Dittmar. 2018. "Universal Molecular Structures in Natural Dissolved Organic Matter." *Nature Communications* 9, no. 1: 3178. <https://doi.org/10.1038/s41467-018-05665-9>.

Zhan, X., H. Bao, J. Niggemann, et al. 2024. "Beyond Hydrology: Exploring the Factors Influencing the Seasonal Variation of the Molecular Composition of Riverine Dissolved Organic Matter." *Journal of Geophysical Research: Biogeosciences* 129, no. 9: e2024JG008014. <https://doi.org/10.1029/2024JG008014>.

Zhi, W., C. Klingler, J. Liu, and L. Li. 2023. "Widespread Deoxygenation in Warming Rivers." *Nature Climate Change* 13, no. 10: 1105–1113. <https://doi.org/10.1038/s41558-023-01793-3>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Environmental variables used in this study. **Table S2:** Molecular traits used in this study. **Figure S1:** The van Krevelen diagram of DOM molecules with significant positive (blue), significant negative (red), and neutral (grey) molecular habitat affinities (iHaf). n is the number of molecules. The dashed lines separating compound classes on the van Krevelen diagram are for visualization only. The compound classes include lipid-like, protein-like, amino sugar-like, carbohydrate-like (Carb), unsaturated hydrocarbon-like (UnsatHC), lignin-like, tannin-like, and condensed aromatic-like (ConHC). **Figure S2:** The distribution of non-significant molecular habitat affinities (iHaf). n is the number of molecules. **Figure S3:** The percentage of molecules for each compound class along the continuum of molecular habitat affinity (iHaf) grouped into 550 equally sized windows (i.e., 500 molecules per window). The x -axis denotes the lowest absolute of iHaf for each window. The compound classes include tannin-like, lignin-like, condensed aromatic-like (ConHC), carbohydrate-like (Carb), unsaturated hydrocarbon-like (UnsatHC), lipid-like, amino sugar-like, and protein-like. **Figure S4:** Principal component analysis (PCA) of 12 molecular traits of DOM. (a) Molecular traits in a two-dimensional orthogonal space of PCA. (b, c) The contributions of molecular traits

on the first two dimensions of PCA. Red dashed line represents the average contribution. Abbreviations of molecular traits are detailed in Table S2. **Figure S5:** Relationship between molecular abundance and sediment respiration rates along the continuum of molecular habitat affinity (iHaf) grouped into 550 equally sized windows (i.e., 500 molecules per window). The x -axis denotes the lowest absolute of iHaf for each window. Blue and red dashed lines highlight the cutoff of 100% significantly positive and negative iHaf for the molecules in each window, respectively. **Figure S6:** The continuous changes in modified aromaticity index (AI_{mod}) and double bond equivalents (DBE) along the continuum of molecular habitat affinity (iHaf) grouped into 550 equally sized windows (i.e., 500 molecules per window). The x -axis denotes the lowest absolute of iHaf for each window. Values are presented as the means \pm SE. Blue and red dashed lines highlight the cutoff of 100% significantly positive and negative iHaf for the molecules in each window, respectively. **Figure S7:** Violin plot of compositional-level habitat affinities ($iHaf_{wm}$) of DOM in waters and sediments. Smaller dots are the $iHaf_{wm}$ for individual samples, and black dots within each violin show the average values. **Figure S8:** Variations in habitat affinities of DOM assemblages ($iHaf_{wm}$) with abundance-weighted traits including the modified aromaticity index (AI_{mod}) and double bond equivalents (DBE) in waters and sediments. Solid lines indicate statistically significant ($p \leq 0.05$) relationships. **Figure S9:** The relative importance of environmental variables in influencing the habitat affinities of DOM assemblages ($iHaf_{wm}$) in (a) waters and (b) sediments. The relative importance was estimated by machine learning-based random forest models. The environmental variables are grouped by hydrography, climate, anthropogenic disturbances, land cover and soil properties with different colors. Abbreviations of environmental variables are detailed in Table S1. Insets show the performance of random forest models, with root mean square error (RMSE) and determination coefficient (R^2) for $iHaf_{wm}$ in waters and sediments. **Figure S10:** The spatial distribution of estimated habitat affinities of DOM assemblages ($iHaf_{wm}$) for (a) waters and (b) sediments in global river basins. Source map of river basins was obtained from HydroBASINS level 02. **Figure S11:** The uncertainty of habitat affinities of DOM assemblages ($iHaf_{wm}$) for (a) waters and (b) sediments in the global river network. The uncertainty was estimated as the standard error of individual predictions of 2000 trees in the random forest model. **Figure S12:** The molecular composition at the compound class level (i.e., the percentage of molecules within each compound class) along the continuum of negative to positive molecular habitat affinity (iHaf). The iHaf continuum was constructed using the moving window method with (a) a window size of 300 molecules and a step of 20 molecules, (b) a window size of 700 molecules and a step of 20 molecules, (c) a window size of 500 molecules and a step of 10 molecules, and (d) a window size of 500 molecules and a step of 30 molecules, respectively. The x -axis denotes the lowest absolute iHaf of each window. **Figure S13:** Gradual changes in mass, H/C ratio, O/C ratio, and Gibbs free energy (GFE) along the continuum of molecular habitat affinity (iHaf). The iHaf continuum was constructed using the moving window method with (a) a window size of 300 molecules and a step of 20 molecules, (b) a window size of 700 molecules and a step of 20 molecules, (c) a window size of 500 molecules and a step of 10 molecules, and (d) a window size of 500 molecules and a step of 30 molecules, respectively. The values of molecular traits are presented as the means \pm SE. The x -axis denotes the lowest absolute iHaf of each window.