Chemodiversity of Dissolved Organic Matter Is Governed by Microbial Biogeography in Inland Waters

Wanzhu Li,∥ Na Liu,∥ Jianfeng Li, Baoli Wang,* Xinjie Shi, Xia Liang, Meiling Yang, Sheng Xu, and Cong-Qiang Liu

ABSTRACT: Dissolved organic matter (DOM) is crucial for the carbon biogeochemical cycle and has a close link with microbiome in aquatic ecosystems; however, the causal relationship between DOM and microbial diversity in inland waters is not very clear so far. Therefore, a national survey of China’s inland waters was conducted, and the DOM chemical composition and microbial community composition were determined by Fourier transform ion cyclotron resonance mass spectrometry and high-throughput sequencing to clarify the abovementioned question. Here, we found that DOM chemodiversity was governed by microbial community assembly in inland waters, not vice versa. Under the control of microbial biogeography, DOM chemodiversity showed a clear geographical distribution difference. Water DOM chemodiversity was mainly constrained by bacterial and archaeal community composition, whereas sediment DOM chemodiversity was mainly controlled by eukaryotic and fungal community composition. In addition, the sediment DOM chemical composition was also affected by the interaction of different microbial groups between waters and sediments. The study is the first to clarify the causal relationship and proposes a microbial regulatory mechanism on the geographical distribution pattern of DOM chemodiversity, thus further deepening the understanding of the DOM biogeochemical cycle.

KEYWORDS: dissolved organic matter, chemodiversity, microbial regulatory mechanism, biogeography, inland water

INTRODUCTION

Dissolved organic matter (DOM) plays an important role in the carbon biogeochemical cycle of aquatic ecosystems, including modulating the emission of CO₂ and CH₄ and supplying for microbial lives.¹,² A handful of water contains thousands of DOM molecules, and with the rapid development of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), these DOM molecules have a great chemodiversity, although the isomeric diversity behind each elemental composition is obscure as yet.³⁻⁵ The resulting diversity could be from two processes. One is the generation process of DOM, which mainly includes the microbial degradation of terrestrial and endogenous non-living particulate organic matter (POM), the exudation of metabolites of living aquatic organisms, the lysis of microbial cells caused by virus infection, and photochemical degradation of the high-molecular-weight DOM;⁶⁻⁸ the other is the DOM consumption process, mainly including the microbial utilization, secondary degradation, and mineralization and photo-degradation of labile DOM.⁹ The microbial metabolites are species-specific,¹⁰ and the variation in microbial community composition can thus affect the DOM chemodiversity. On the other hand, labile DOM may serve as a matter and energy source of microorganisms. Since some microorganisms may prefer a specific niche including a certain group of DOM, change in DOM chemical composition is considered to be an important factor affecting microbial community assembly.¹¹,¹² As such, microbial and DOM diversity is closely linked, and it has been reported that the co-variation in microbial and DOM diversity can influence the functioning of aquatic ecosystems.¹³⁻¹⁵ However, it is not clear whether the microbial diversity controls the DOM chemical diversity or vice versa.

Nowadays, the understanding of DOM chemodiversity and its regulatory mechanism in inland waters is still in the infancy.⁵,¹⁶ The environmental factors such as hydrology and nutrient availability that may affect DOM production and consumption could modulate DOM chemodiversity.⁵,¹⁶⁻¹⁸ Furthermore, the DOM concentration in the pore waters of sediments is much higher than that in the overlying waters due
to more degradation of POM, and then the DOM releases, resulting in the more DOM complexity in the upper waters.\textsuperscript{19} Surprisingly, it is found that DOM chemical composition has an obvious geographical distribution difference either in water or in sediment,\textsuperscript{14,20,21} of which the relevant control mechanism remains to be clarified. Similarly, for the aquatic microorganisms in inland waters, their community composition has a clear geographical difference as well, which is shaped by environmental filtering and biological interaction (i.e., deterministic processes based on niche theory) and by some random events such as drift and dispersal (i.e., stochastic processes based on neutral theory).\textsuperscript{22−24} Some statistical methods for studying microbial assembly have been used to understand the forming mechanism of DOM chemodiversity.\textsuperscript{25} Considering that microorganisms drive the DOM biogeochemical cycle in aquatic ecosystems,\textsuperscript{26,27} we hypothesized that the microbial community assembly governs the DOM chemodiversity in inland waters.

A national survey of inland waters was conducted, and we investigated the waters, sediments, and related environmental factors to verify the abovementioned hypothesis. The FT-ICR MS and high-throughput sequencing were respectively used to analyze DOM-detailed molecular information and microbial community composition (including bacteria, archaea, eukaryote, and fungi). The causal relationship between DOM chemodiversity and microbial community diversity was analyzed by using multiple statistical methods. The main aims of this study are as follows: (1) to understand the geographical distribution and regulatory factors of DOM chemical composition under a large environmental gradient; (2) to discern the geographical distribution pattern and influencing factors of microbial community structure; and (3) to explore the regulatory mechanism of microbial community composition on DOM chemodiversity in inland waters. Determining the causal relationship between DOM chemical and microbial community diversity is crucial to clarify the microbial driving mechanism in the DOM biogeochemical cycle. Therefore, this study will deepen the understanding of DOM biogeochemistry.

### MATERIALS AND METHODS

**Study Area and Sampling.** A national survey of 35 inland waters (including wetlands, lakes, and reservoirs) was conducted in July and August 2021 with a range of 24.19–45.46°N, 99.39−126.35°E in China (Figure 1). The 36 surface water samples and 11 surface sediment samples were collected from the lakeshore (about 10 m from the shore). The detailed information of sampling sites is given in Table S1. Sediment columns were collected using a rod holding a shallow water columnar sediment sampler, and then the surface sediments (surface 1 cm) were collected using sterile polyethylene bags and stored at −20°C for subsequent parameter analysis and DNA extraction. Water temperature (WT), pH, and dissolved oxygen concentration (DO) were determined in situ by an automated multiparameter profiler (YSI, EXO1, USA) with precorrection. Effective quantum yield (yield) and chlorophyll a concentration (Chl a) were measured by a phytoplankton analyzer (Phyto-PAM, Walz, Germany).

![Figure 1. Map of the studied inland waters in China. The name and detailed sampling sites are given in Table S1.](image-url)
water samples was filtered through a 0.7 μm Whatman GF/F glass fiber membrane that was pre-combusted (450 °C for 8 h), and the filtered waters were collected in 60 mL brown glass bottles and 1.5 L polyethylene bottles for the determination of dissolved organic carbon (DOC) concentration and DOM chemical composition, respectively. They were stored at 4 °C in the dark. DOM samples were acidified with 1 mol L−1 HCl to pH = 2. The microbial samples were collected using 0.22 μm Millipore cellulose acetate membrane for DNA extraction, and the filtered waters were collected in 15 mL centrifuge tube (ThermoFisher, USA) and 60 mL polyethylene bottle, respectively, and stored at 4 °C for the determination of cation and anion concentrations and nutrient concentrations. All samples were treated within 12 h after sampling as soon as possible.

The 0.5 g freeze-dried sediments were placed in a 50 mL centrifuge tube (ThermoFisher, USA) with 50 mL Milli-Q ultra-pure waters and shaken using 170 rpm for 8 h at 25 °C. The samples were centrifugated at 4600 g for 15 min, and the supernatant was passed through the 0.45 μm cellulose acetate membrane. The filtered waters were the sediment leaching solutions used to prepare to measure DOC concentration and DOM chemical composition, and the DOM samples were acidified with 1 mol L−1 HCl to pH = 2.

Physical and Chemical Parameter Measurement. Nutrient concentrations, including nitrate nitrogen (NO−3-N), nitrite nitrogen (NO−2-N), ammonia nitrogen (NH4-N), phosphate phosphorus (PO4-P), and dissolved silicon (Dsi) concentrations, were measured by Skalar (SAN++, the Netherlands) with detection limits of 0.01, 0.005, 0.02, 0.02, and 0.02 mg L−1, respectively. The concentration of dissolved inorganic nitrogen (DIN) was the sum of NO−3-N, NO−2-N, and NH4-N concentrations. The concentration of dissolved inorganic phosphorus (DIP) was referred to PO4-P concentration. Cation and anion concentrations (i.e., Na+, K+, Mg2+, Ca2+, Cl−, SO42−, and NO−3) were determined by ion chromatography (ICS-5000, Thermo Fisher, USA) with detection limits of 0.1 mg L−1. Total alkalinity (ALK) was obtained by titration with standard hydrochloric acid. The dissociation constants for carbonic acid were calculated by the following formula:

\[ K_a = \frac{X_1}{X_2} \]

where \( X_1 \) (%) was the relative abundance of a certain DOM chemical component, \( A \) was the sum of the peak intensity of a certain DOM chemical component, and \( A_i \) was the sum of the peak intensities of all DOM chemical components in the sample.

DNA Extraction and Sequencing. The total DNA of sample was extracted by the E.Z.N.A TM Water DNA Kit (OMEGA, USA) according to manufacturer’s protocols. The concentration and purity of DNA were determined by Multiskan GO (Thermo Fisher, USA). The details for DNA extraction are given in Wang et al.34 High-throughput sequencing was performed on Illumina NovaSeq 6000 platform at Meige Technology Co., Ltd., Guangzhou, China. The S15F (GTGCCACGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) were used as primers for bacteria. The Arch340F (CCCTAYGGGGGCASCAG), Arch1000F (GGCCATGCAYWCYTCTC), Uni519F (CAGYMGCCRCGGKAAHACC), and Arch806R (GGAC-TACNCGGTTTCTAAT) were used as primers for archaea. The S56F (CCAGCASCYGCGGTAATTCC) and 981R (ACCTTCTGTCTTTGATYRTAGA) were used as primers for eukaryote. The BD-ITS1F (CTTGTGTCAGAGGAA-TAA) and BD-ITS2 (GCTGGTTTTCCTCTATGATGC) were used as primers for fungi. Raw fastq data were quality-filtered using Trimmomatic35 to remove contaminating adaptors, low quality ends of reads, and short length reads (<200 bp). The overlapping paired-end reads with sequence mismatching <5 bp and alignment similarity >90% were merged using FLASH.36 Operational taxonomic units (OTUs) were clustered in UPARSE software at 97% consistency level.37 Singleton and doubleton OTUs representing sequencing errors were removed, and the representative OTUs with the highest occurrence frequency were assigned using Silva, RDPII, and Greengene.38 To equalize sequencing depth, each sample was rarefied to the minimum sequencing depth, and sequence normalization was performed using MOTHUR v.1.33.3.39

Data Analysis. According to the stoichiometric ratio of detected molecular formula (H/C and O/C) and element combination (C, H, O, N, P, and S), two DOM classification methods were determined referring to Kellerman et al.5 Corresponding to the compounds in natural organic matter, the compound classes based on the Van Krevelen diagram were classified: lipid (0 < O/C < 0.3, 1.5 < H/C < 2.0), protein (0.3 < O/C < 0.55, 1.5 < H/C < 2.2), amino sugars...
Figure 2. Compositions of compound class, elemental combination, and microbial community in the water (a–c) and sediment (d–f) samples. The sampling sites were arranged from low to high longitude (color bar). Microbial community: Bac, bacteria; Arc, archaea; Euk, eukaryote (excluding fungi); Fun, fungi. Elemental combinations for DOM composition: CHO, CHON, CHOS, CHOP, CHONS, and others. Compound classes for DOM composition: Lip, lipid; Pro, protein; Ami, amino sugars; Car, carbohydrates; Hyd, unsaturated hydrocarbons; Lig, lignin-like; Aro, condensed aromatics; and Tan, tannin. The detailed information of sampling sites is given in Table S1.

(0.5 < O/C < 0.67, 1.5 < H/C < 2.2), carbohydrates (0.67 < O/C < 1.2, 1.5 < H/C < 2), unsaturated hydrocarbons (0 < O/C < 0.1, 0.7 < H/C < 1.5), lignin-like (0.1 < O/C < 0.67, 0.7 < H/C < 1.5), condensed aromatics (0 < O/C < 0.67, 0.2 < H/C < 0.7), and tannin (0.67 < O/C < 1.2, 0.5 < H/C < 1.5) (see Figure S1). The other is elemental combination, including CHO, CHON, CHOS, CHOP, and CHONS compounds. Double bond equivalent (DBE) and its related
parameters were utilized to indicate the unsaturation of DOM. DBE minus oxygen (DBE-O) can describe unsaturation by excluding all the possible C=O bonds in the −COOH functional group. The DBE to carbon ratio (DBE/C) can reveal aromatic or condensed aromatic structures in high density of C=C bonds.

The Van Krevelen diagram and microbial community composition and geographical distribution charts were plotted by Origin 2021. The t test was used to determine significant differences between groups within the 95% confidence interval by IBM SPSS Statistics 23. Solar radiation data for each sampling site were obtained from agrometeorological data (www.wheata.cn). Statistical analysis was performed by R software (version 4.2.0). The Bray–Curtis distance of microbial community composition and Euclidean distance of DOM composition were calculated using the R software “vegan” package for the characterization of their composition dissimilarities between the sampling sites. Spearman’s correlation analyses were performed using “corrplot” package for understanding the correlation relationship between the variables. Random forest analyses were performed using “randomForest” and “rfPermute” package to discern the significant predictors for a given parameter. The structural equation model (SEM) was established using “lavaan” package to test the direct and indirect relationships between the variables, validate different conceptual models, and select the optimal model. The Shannon–Wiener index (H') was calculated using the “picante” package to represent DOM and microbial α-diversity. DOM molecular chemodiversity (H’-Mol) was calculated using the molecular formula normalized peak intensity. DOM elemental combination diversity (H’-Ele) and DOM compound class diversity (H’-Com) were calculated by their relative abundance. The diversities of bacteria (H’-Bac), archaea (H’-Arc), eukaryotes (H’-Euk, excluding fungi), and fungi (H’-Fun) were calculated using their OTUs. Microbial diversity (H’-Mic) was weighted calculated using the relative abundance of the abovementioned four groups; the calculated equation was as follows:

$$ H'_{\text{Mic}} = f_B \sum_{i=1}^{n} (P_B \ln P_B) + f_A \sum_{i=1}^{n} (P_A \ln P_A) + f_E \sum_{i=1}^{n} (P_E \ln P_E) + f_F \sum_{i=1}^{n} (P_F \ln P_F) \quad (i = 1, 2, 3...n) $$

$$ f_B + f_A + f_E + f_F = 1 $$

where f is the proportion of OTU abundance of a certain group (B, bacteria; A, archaea; E, eukaryote; F, fungi) to the total OTU abundance of the microbial community, P is the proportion of each OTU within a certain group to the total.
OTU of that group, and \( n \) is the number of OTUs in a certain group.

\section*{RESULTS}

\textbf{DOM Composition and Related Environmental Parameters.} A total of 11,453 molecular formulas and 8661 molecular formulas were identified by FT-ICR MS in the inland waters and sediments, respectively. The sum of normalized peak intensities was used to calculate the species accumulation and rank abundance curves. 17 of the 36 water samples and 4 of the 11 sediment samples had already accumulated the 95\% of their respective total molecular number, and the chemical molecules with normalized intensities in the highest 40\% were in at least 70\% of the corresponding water and sediment samples (Figure S2), indicating that the samples are sufficient and representative.\textsuperscript{5}

The DOM average molecular formula based on the peak-\textit{intensity}-\textit{weighted} calculation was \( C_{16.0}.H_{19.41}.O_{7.78}.N_{0.209}.S_{0.19}.P_{0.02} \) in the waters and \( C_{14.88}.H_{21.04}.O_{6.75}.N_{0.23}.S_{0.30}.P_{0.03} \) in the sediments (Tables S2 and S3). The highest relative abundance in the waters and sediments were lignin-like (70.83 and 35.50\%, respectively) for the compound classes and CHO compounds (67.08 and 45.03\%, respectively) for the elemental combinations (Figure 2). The range and average of DOM’s DBE were both greater in the waters (6.19–9.04, 7.56) than in the sediments (4.16–6.95, 5.58; Tables S2 and S3). Lignin-like had no significant correlation with either DBE-\( \Omega \) (\( r = 0.01, p > 0.05 \)) or DBE/\( \Omega _{wa} \) (\( r = 0.06, p > 0.05 \)) in the waters, respectively (Table S2). However, in the sediments, lignin-like was significantly positively correlated with both DBE-\( \Omega _{wa} \) (\( r = 0.50, p < 0.05 \)) and DBE/\( \Omega _{wa} \) (\( r = 0.97, p < 0.001 \)), respectively (Table S3). In the waters, H’-Mol ranged from 6.26 to 7.92, with an average of 7.37; H’-Com and H’-Ele ranged from 0.59 to 1.29 and 0.32 to 1.23, with an average of 1.01 and 0.94, respectively. Their variation ranges in the sediments were smaller than these in the waters, but their averages in the former were larger than those in the latter (Tables S2 and S3). All three of the abovementioned DOM chemodiversity parameters had a similar spatial distribution characteristic in both waters and sediments, and the H’-Mol was peaked at the range of 24–28\°N, 100–108\°E and 31–40\°N, 110–120\°E (Figures 3a and S3).

In the waters, the average of WT, pH, DO concentration, and Chl a concentration was 26.19 \textdegree C, 8.45, 9.64 mg L\(^{-1}\), and 42.53 \( \mu \)g L\(^{-1}\), respectively, showing a large variation (Table S2), pH and DO were significantly correlated (Figure S4). The average of CO\(_2\), DIN, DIP, and DSI concentrations was 43.05, 70.77, 2.63, and 2.06 \( \mu \)mol L\(^{-1}\), respectively (Table S2). DOC concentration ranged from 1.16 to 15.31 mg L\(^{-1}\), with an average of 4.88 mg L\(^{-1}\), and was significantly correlated with longitude (Figure S4). H’-Mol was significantly correlated with WT and yield, H’-Com was significantly correlated with latitude and longitude, and H’-Ele was significantly correlated with latitude, WT, and Chl a concentration, respectively (Figure S4). In the sediments, DOC concentration ranged from 0.92 to 15.72 mg L\(^{-1}\), with an average of 4.29 mg L\(^{-1}\). The average of MC and PO was 43.77 and 61.44\%, respectively (Table S3). DOC concentration, MC, and PO were significantly correlated with each other. Both H’-Mol and H’-Ele were significantly correlated with DOC concentration (Figure S4).

\textbf{Microbial Composition.} In the waters, a total of 12,405 OTUs of planktonic bacteria were obtained by high-throughput sequencing and clustered into 49 phyla, and the top two phyla were Proteobacteria and Actinobacteria; a total of 7664 OTUs of planktonic archaea were obtained and clustered into 9 phyla, and the top two phyla were Nanoarcheaeota and Euryarchaeota; a total of 4942 OTUs of eukaryotes were obtained and clustered into 61 phyla, and the top two phyla were Ochrophyta and Unassigned; a total of 11,814 OTUs of fungi were obtained and clustered into 26 phyla, and the top two phyla were Unassigned and Ascomycota. In the sediments, a total of 19,344 OTUs of bacteria were obtained and clustered into 72 phyla, and the top two phyla were Proteobacteria and Bacteroidetes; a total of 12,660 OTUs of archaea were obtained and clustered into 64 phyla, and the top two phyla were Nanoarchaeota and Crenarchaeota; a total of 2775 OTUs of eukaryotes were obtained and clustered into 64 phyla, and the top two phyla were Ochrophyta and Euryarchaeota. In the sediments, bacteria were the highest relative abundance microbial community (Figure 2). All four microbial groups had different spatial community structure (Figures S5 and S6).

The H’-Mic ranged from 2.65 to 4.94 and from 3.40 to 4.88, with an average of 3.96 and 4.17 in the waters and sediments, respectively (Tables S2 and S3) and showed a lower value in the range of 24–30\°N, 103–112\°E and a higher value in the range of 25–31\°N, 110–120\°E (Figures 3a and S3). The averages of H’-Mic, H’-Bac, and H’-Arc in the waters were lower than those in the sediment, while both had similar H’- Fun and H’-Euk averages (Tables S2 and S3). Bacterial and archaeal community diversity showed a similar spatial distribution pattern in the waters and sediments. Nevertheless, the spatial distribution patterns of eukaryotic and fungal community diversity were special (Figure S7). H’-Bac and H’-Arc in the waters were significantly positively correlated with latitude, while H’-Euk and H’-Fun in the sediments had a significantly negative correlation with longitude (Figure S4). In the waters, H’-Mic, H’-Arc, and H’-Bac were significantly correlated with DSI concentration. H’-Euk was significantly correlated with DIP concentration, and H’-Fun was significantly correlated with pH and DOC concentration, respectively (Figure S4a). In the sediments, both H’-Mic and H’-Fun were correlated with MC and PO, respectively (Figure S4b).

\textbf{Relationship between DOM and Microbial Composition.} In the waters, dissimilarity of DOM composition and microbial community increased significantly with increasing geographic distance, but in the sediments, only the latter was found the distance-decay relationship (Figure 3b,c). DOM molecular composition was significantly correlated with microbial community diversity (Figure 4). In the waters, microbial diversity was mostly negatively correlated with DOM molecules when H/C \( \geq 1.5 \) (Figure 4), and \( f_{wa} \) was significantly positively correlated with H’-Com (Figure S8a). In the sediments, although there were few DOM molecules correlated with microbial diversity, they all had high correlations; DOM molecules showed different correlations with archaeal community diversity from eukaryotic community diversity (Figure 4), and H’-Bac and H’-Com were significantly positively correlated (Figure S8b). In addition, the dominant microbial phyla were highly correlated with water DOM composition. For example, the dominant phyla of bacteria and...
Factors Affecting Microbial and DOM Compositions in Inland Waters. Microbial community plays a key role in maintaining ecological function of aquatic ecosystems, and a distance–decay relationship is the basis for understanding microbial biogeographical distribution pattern. In this study, aquatic microbial community composition had an obvious distribution difference on a large environmental gradient and spatial scale (about a 3500-km distance): an increase in the difference with increasing geographic distance (Figure 3b,c), resulting in a significant correlation between microbial community $\alpha$-diversity (i.e., $H^\prime$) and latitude (and/or longitude) (Figure S4). Overall, environmental variables (e.g., $pH$, $CO_2$, and $DSi$) shape the microbial community structure; however, different microbial groups are sensitive to different environmental factors. For example, bacterial community composition could be affected by solar radiation and $DSi$, whereas fungi community composition could be affected by DOC. Considering that variations in inorganic nutrient concentrations (e.g., $DSi$ and DIP) can also reflect their assimilation by phytoplankton, the interaction between phytoplankton and other microbial community may constrain the overall microbial community assembly. In addition, it is well known that the $pH$, DO, and solar radiation determine the water acidity, redox condition, and available light energy in microbial habitat, and thus they are the important influencing factors in microbial community assembly.

**DOM chemical composition will be determined by their sources and subsequent processing. Lignin is an important organic composition of terrestrial vascular plants and an ideal biomarker for terrestrial organic matter. In the sediments rather than the waters, lignin being the main DOM component means that terrestrial plant debris moved into water is stored in the sediment after continuous degradation, as lignin is generally difficult to be degraded by microorganisms.**

Phytoplankton are an important primary producer assimilating inorganic nutrients and can release DOM through cell autolysis and/or exudation of metabolites. An increase in WT can enhance phytoplankton growth, and phytoplankton consume $CO_2$ and release $O_2$ by photosynthesis, resulting in an increase in water $pH$ and DO. The tight relationship between $pH$ and DO suggests that phytoplankton photosynthesis is domimative. Therefore, WT and phytoplankton photosynthetic efficiency show a close relationship with DOM molecular composition (Figure S4a). It has been reported that phytoplankton blooming induced by excessive nutrient input can enhance the generation and transformation of DOM and then promote the DOM chemodiversity. Furthermore, climate and hydrology are the important influencing factors on DOM chemodiversity. For example, increasing water retention time can enhance both the in-lake production of aliphatic compounds and degradation of imported unsaturated, polyphenolic, and condensed poliaromatic DOM. DOC concentration of the sediments was relevant to porosity and MC (Figure S4b), and this situation generally occurs in lake and river sediments. The high MC is associated with microbial activity and then promotes the transformation from organic

**Discussion**

Microbial communities were interconnected between the waters and sediments (Figure S10), and these interactions ultimately influenced DOM chemodiversity in the sediments (Figure S5g).

**Figure 4.** Comparison of DOM composition patterns driven by microbial community composition in the water and sediment samples. The color scale indicates the strength and direction of the significant molecular-specific Spearman correlations with microbial community diversity ($p < 0.05$). Microbial community composition at the OTU level: Mic, total microbial community; Bac, bacteria; Arc, archaea; Euk, eukaryote (excluding fungi); Fun, fungi. Compound classes based on the Van Krevelen diagram: lipid (I), protein (II), amino sugars (III), carbohydrates (IV), unsaturated hydrocarbons (V), lignin-like (VI), tannin (VII), and condensed aromatics (VII) (more detailed information is given in Figure S1).
matter into hydrated organic matter. Protein and CHOS compounds were found more in the sediment than in the waters. Protein with high H/C is usually autochthonous, and the DOM of microbial source will be enriched more with heteroatom S-containing compounds.

In the recent studies of China’s coastal wetlands, aliphatic compounds (relative abundance ranging from 15.6 to 70.8%) and CHOS compounds (relative abundance ranging from 8.5 to 90.6%) have been reported as the important DOM components in the sediments. Aliphatic compounds are usually microbial source, and S-containing proportion is high. These all suggest that DOM has various sources and transformation processes.

Mechanism of Microbial Control on DOM Chemo-diversity in Inland Waters. In this study, we found that there was no control of DOM chemodiversity on the microbial community assembly, but the latter shaped the former (Figure 5); this finally resulted in an obvious geographical distribution difference in DOM chemical diversity (Figures 3a and S3). Aquatic microbial metabolites and the assimilation of DOM are species-specific, and microbial community assembly that drives the production and consumption of DOM thus becomes a critical factor determining DOM chemodiversity. The significant negative correlation between bacteria (or archaea) and labile DOM (i.e., H/C > 1.5, including lipid, protein, amino sugar, and carbohydrate) indicated that the former has a strong utilization of the latter, whereas microeukaryotes prefer refractory DOM (i.e., H/C < 1.5, including lignin-like and tannin) (Figure 4). Specifically, Proteobacteria and Thau- marchaeota constrain the conflict direction between lipid and tannin, whereas Bacteroidetes and Nanoarchaeota jointly regulate the cooperation evolution of CHON and CHOP compounds (Figure 5d). Overall, microbial activity consumes the labile DOM and increases the relative abundance of lignin-like and CHO compounds, and specific microbial community composition creates specific DOM composition (Figures 4 and 5). For example, increasing bacterial relative abundance can enhance the diversity of DOM compound class (Figure S8a).

Compared with in the waters, the organic matter is continuously accumulated and stored for a longer time in the sediments, where the DOM chemodiversity is affected not only by eukaryotic and fungal community assembly but also by their interaction between waters and sediments (Figures 5g, S9, and S10). It has been reported that Cryptomycota is a

Figure 5. (a) Spearman correlation between microbial community and DOM in the water samples. The numbers represent significant correlation coefficients (p < 0.05). Pink, blue, green, and yellow represent eukaryote (Euk), archaea (Arc), bacteria (Bac), and fungi (Fun), respectively. The abbreviations for DOM compositions are given above. Random forest analyses for the influencing factors for microbial community diversity (b), DOM composition diversity (c), and DOM composition coexistence (d) in the water samples. The numbers represent significant mean squared error (p < 0.05), with higher mean squared error implying more important predictors, and accuracy important measure was computed for each tree and average over the whole forest (500 trees). Blue and red indicate the positive and negative relationship, respectively, and the color shade indicates the correlation strength in (a–d). Structural equation modeling describing the direction relationships among the relevant parameters in the water (blue font) and sediment (orange font) samples (e–g). Numbers adjacent to arrows indicate the size and direction of relationship effect. RMSEA, root mean square error of estimation; p, p value for a test of close fit, CFI, comparative fit index. ** and * indicate the significance levels of 0.01 and 0.05, respectively. The other abbreviations are given in the text.
phylogenetically diverse but ecologically mysterious group of early-diverging, fungi or fungal-like eukaryotes, pervasive in freshwater sediment, and probably influences ecosystems significantly through control of primary producer and consumer. This study found that it was an important phylum for microeukaryote linking the sediments with the upper waters. Furthermore, some bacteria can degrade lignin, and in this study, Actinobacteria was found for the lignin degradation in the waters (Figure 5a).

SEM analysis further demonstrated that bacterial, archaeal, eukaryotic, and fungal community and their interaction govern the formation of DOM chemodiversity (Figure 5e–g). Furthermore, the ratio of $H^\prime$-Mol to $H^\prime$-Mic for each sampling site had no significant deviations from their average ratio (t test, $p = 1$), indicating that the constraint of microorganisms on DOM chemodiversity will be popular at a large geographical scale. All these demonstrated that DOM chemodiversity is governed by microbial biogeography in inland waters, not vice versa. The opinion that DOM composition affects microbial community diversity can be seen from the studies of the interaction between DOM and bacterial community composition, and the explanation is based on the fact that specific niche formed by specific DOM composition benefits the specific bacterial community assembly. However, our study does not support this, perhaps in part because there are complex interactions among bacteria, archaea, and microeukaryote in water. As such, when clarifying the causal relationship between planktonic microbial and DOM diversity, the biological effects of all microbial groups should be considered. This is a great challenge, and culture experiments in the field and/or laboratory will provide some direct evidence. In addition, in this study, although there was a geographical distribution difference of DOM chemodiversity, the distance-decay relationship of DOM composition in the sediments was not significant, probably as the number of sediment samples was relatively smaller than that of water samples.

This study is the first to clarify the causal relationship between microbial and DOM diversity in inland waters. In general, there are significant geographical distribution differences for planktonic bacterial, archaeal, eukaryotic, and fungal community composition and DOM chemical composition. Water DOM chemodiversity was mainly constrained by bacterial and archaeal community composition, whereas sediment DOM chemodiversity was mainly controlled by eukaryotic and fungal community composition. In addition, the sediment DOM chemical composition was also affected by the interaction of different microbial groups between waters and sediments. This study not only provides an idea to explain the formation mechanism of DOM complexity caused by the connectivity between waters and sediments but also proposes a microbial regulatory mechanism on the geographical distribution pattern of DOM chemical composition, thus further deepening the understanding of the DOM biogeochemical cycle.

**ASSOCIATED CONTENT**

**Data Availability Statement**

Bacterial, archaeal, eukaryotic, and fungal raw data were deposited in the NCBI SRA database with the accession numbers of PRJNA907153. FT-ICR MS data are referred to the Supporting excel file.

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c00896. Additional experimental details of FT-ICR MS sample preparation for sediment DOM, detailed information of sampling sites, range and average for each physical and chemical parameter of water and sediment samples, chemical components of water and sediment DOM in China’s inland waters, molecular distributions of FT-ICR MS detected compounds in China’s inland waters, geographical distributions of microbial community diversity and DOM chemodiversity in the water and sediment samples, Spearman correlation among environmental factors, microbial community composition, and DOM composition for the water and sediment samples, community compositions of eukaryote, archaea, bacteria, and fungi in the water and sediment samples, geographical distribution of community diversity of bacteria, archaea, eukaryote, and fungi in the water and sediment samples, bacterial relative abundance versus compound class diversity in the water samples, bacterial community diversity versus compound class diversity in the sediment samples, Spearman correlation between microbial community composition and DOM composition in the sediment samples, and Spearman correlation of microbial community compositions between the water and sediment samples (PDF) FT-ICR MS data (xlsx)

**AUTHOR INFORMATION**

**Corresponding Author**

Baoli Wang — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China; Tianjin Bohai Rim Coastal Earth Critical Zone National Observation and Research Station, Tianjin 300072, China; orcid.org/0000-0003-1437-5695; Phone: +86 (0)22 27405053; Email: baoli.wang@tju.edu.cn; Fax: +86 (0)22 27405051

**Authors**

Wanzhu Li — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Na Liu — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Jianfeng Li — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Xinjie Shi — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Xia Liang — State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China; orcid.org/0000-0003-3882-5406

Meling Yang — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Sheng Xu — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

Cong-Qiang Liu — Institute of Surface-Earth System Science, School of Earth System Science, Tianjin University, Tianjin 300072, China

**Funding**

This work was supported by the National Natural Science Foundation of China (Grant No. 41977300), the Ministry of Science and Technology of China (Grant No. 2017FY110800), the National Key Research and Development Program of China (Grant No. 2016YFA0602702), the Major Scientific Research Project of the Chinese Academy of Sciences (Grant No. XDA19030200), and the China Postdoctoral Science Foundation (Grant No. 2019M652873).
Author Contributions
W.L. and N.L. contributed equally to this work. B.W. developed the idea, B.W., S.X., and C.Q.L. optimized the study, W.L., N.L., J.L., X.S., X.L., and M.Y. performed the experiment and field work, W.L., N.L., and B.W. produced the first draft, and all co-authors contributed to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS
This work was financially supported by the National Natural Science Foundation of China (42293264), the National Science & Technology Fundamental Resources Investigation Program of China (2021FY101000), and the Haihe Laboratory of Sustainable Chemical Transformations and Tianjin Municipal Science and Technology Bureau (22ZYJDJC00140). The authors are grateful to the colleagues for their help in the field and the reviewers for their valuable comments.

REFERENCES

Drivers of bacterial 


Drivers of bacterial 