

# Ecological networks of dissolved organic matter and microorganisms under global change

Ang Hu<sup>1,2</sup>, Mira Choi<sup>3</sup>, Andrew J. Tanentzap<sup>4</sup>, Jinfu Liu<sup>1,5</sup>, Kyoung-Soon Jang<sup>3</sup>, Jay T. Lennon<sup>6</sup>, Yongqin Liu<sup>7,8</sup>, Janne Soininen<sup>9</sup>, Xiancai Lu<sup>10</sup>, Yunlin Zhang<sup>1</sup>, Ji Shen<sup>11</sup> & Jianjun Wang<sup>1,12</sup>✉

Microbes regulate the composition and turnover of organic matter. Here we developed a framework called Energy-Diversity-Trait integrative Analysis to quantify how dissolved organic matter and microbes interact along global change drivers of temperature and nutrient enrichment. Negative and positive interactions suggest decomposition and production processes of organic matter, respectively. We applied this framework to manipulative field experiments on mountainsides in subarctic and subtropical climates. In both climates, negative interactions of bipartite networks were more specialized than positive interactions, showing fewer interactions between chemical molecules and bacterial taxa. Nutrient enrichment promoted specialization of positive interactions, but decreased specialization of negative interactions, indicating that organic matter was more vulnerable to decomposition by a greater range of bacteria, particularly at warmer temperatures in the subtropical climate. These two global change drivers influenced specialization of negative interactions most strongly via molecular traits, while molecular traits and bacterial diversity similarly affected specialization of positive interactions.

<sup>1</sup> State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China. <sup>2</sup> College of Resources and Environment, Hunan Agricultural University, Changsha 410128, China. <sup>3</sup> Bio-Chemical Analysis Team, Korea Basic Science Institute, Cheongju 28119, South Korea. <sup>4</sup> Ecosystems and Global Change Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK. <sup>5</sup> Nanchang Institute of Technology, Nanchang 330099, China. <sup>6</sup> Department of Biology, Indiana University, Bloomington, IN 47405, USA. <sup>7</sup> Center for the Pan-third Pole Environment, Lanzhou University, Lanzhou 730000, China. <sup>8</sup> State Key Laboratory of Tibetan Plateau Earth System, Resources and Environment, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China. <sup>9</sup> Department of Geosciences and Geography, University of Helsinki, Helsinki FIN-00014, Finland. <sup>10</sup> State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210093, China. <sup>11</sup> School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China. <sup>12</sup> University of Chinese Academy of Sciences, Beijing 100049, China. ✉email: [jjwang@niglas.ac.cn](mailto:jjwang@niglas.ac.cn)

**D**issolved organic matter (DOM) is one of the largest pools of carbon in aquatic ecosystems<sup>1</sup> and its fate is intimately linked with the metabolism of complex microbial communities<sup>2</sup>. Microbial consortia moderate the diversity of molecules in DOM by degrading larger molecules into smaller molecules and by synthesizing more refractory compounds from labile substrates<sup>3,4</sup>. Together, these metabolic processes lead to the emergence of molecular traits (Box 1), such as molecular weight, chemical structure, stoichiometry, oxidation state, and bioavailability<sup>5–7</sup>, all of which have consequences for the fate and persistence of DOM<sup>8,9</sup>. Because it serves as an energy and carbon source for metabolism, DOM also influences the diversity, structure, and function of microbial communities<sup>10–12</sup>. The resulting associations between DOM and microbes can now be characterised in both aquatic<sup>13–15</sup> and terrestrial<sup>16</sup> ecosystems owing to recent advances in ultrahigh-resolution mass spectrometry and high-throughput sequencing. Ultrahigh-resolution mass spectrometry specifically identifies individual molecular formulae within DOM pools. This approach provides more information on the diversity and traits of DOM than available bulk measurements, like those based on absorbance and fluorescence spectroscopy, which are generally lacking in resolution. Despite these technological developments, challenges remain about how DOM-microbe associations can be quantified in nature, and are interactively and independently affected by global environmental change through human-induced perturbations leading to elevated temperatures and nutrient enrichment.

The effects of global change on DOM-microbe associations can be viewed through three proximal drivers (Fig. 1). First, energy supply, such as primary productivity, represents the major input of DOM that fuels microbial metabolism<sup>17,18</sup>. Elevated temperature and nutrient inputs often stimulate primary productivity in ways that influence the composition and availability of organic matter<sup>19,20</sup>. For instance, microbial decomposition of DOM can be limited by substrate and nutrient availability that provide sufficient energy and material for synthesis of requisite extracellular and transport enzymes<sup>21,22</sup>. Second, DOM-microbe associations are affected by diversity. For example, an increase in the diversity of DOM can promote microbial diversity, and vice versa<sup>14</sup>. Such patterns may arise because resource diversity promotes microbial specialization during biochemical transformations by creating

more unique resource niches for consumers to partition<sup>23,24</sup>. Likewise, higher microbial diversity can provide more metabolic pathways (but not always elevated activity) to decompose and produce molecules, which can influence the vulnerability of DOM to degradation<sup>3</sup>. Third, DOM-microbe associations depend on the molecular traits of DOM, such as its bioavailability, measured by chemical proxies such as H/C ratios of individual molecules<sup>25</sup>, as well as life-history traits of microbes (e.g., *r*-selected copiotrophs versus *K*-selected oligotrophs<sup>26</sup> or resource generalists versus resource specialists<sup>24</sup>).

To examine how DOM-microbe associations vary under global change because of the three aforementioned proximal drivers, we developed a framework called Energy-Diversity-Trait integrative Analysis (EDTiA) (Fig. 1). The first step of EDTiA involves the construction of bipartite networks<sup>27</sup> to quantify the degree of specialization between organic molecules and microbial taxa (Box 1). In the DOM-microbe networks, individual DOM molecules are only connected to microbial taxa that use that specific molecule, while the direct interactions within molecules or taxa are not explicitly considered. The second step of EDTiA involves the investigation of ecological networks using an interaction specialization metric,  $H_2'$ , which is derived from Shannon index<sup>28</sup> (Box 1). By taking resource-consumer relationships among a collection of samples into account, an elevated  $H_2'$  means that there is a high degree of specialization between DOM and microbes<sup>28</sup>, where in the extreme example, one bacterial taxon consumes or produces a single DOM molecule. By contrast, lower  $H_2'$  values reflect a more generalized bipartite network where different DOM molecules can be used by a large number of bacterial taxa. The third step of EDTiA provides a statistical-based means for assessing the relative importance of global change on the specialization of DOM-microbe associations via the three proximal drivers (Fig. 2).

In this study, we developed and used the EDTiA framework to evaluate how DOM-microbe associations were independently and jointly influenced by temperature and nutrient enrichment (Figs. 1, 2) in a manipulative field experiment on two mountainsides, a subtropical one in China and a subarctic one in Norway<sup>29</sup>. This approach involved creating microcosms with consistent initial DOM composition but different locally colonised microbial communities and newly produced endogenous

### Box 1 | Glossary of terms for merging organic geochemistry and ecology

**Molecular composition:** The chemical composition of dissolved organic matter (DOM) that is a complex mixture of reduced carbon compounds bound to heteroatoms such as oxygen, nitrogen, phosphorus, and sulphur. The composition is characterised by the identity and intensity of molecular formulae in a sample.

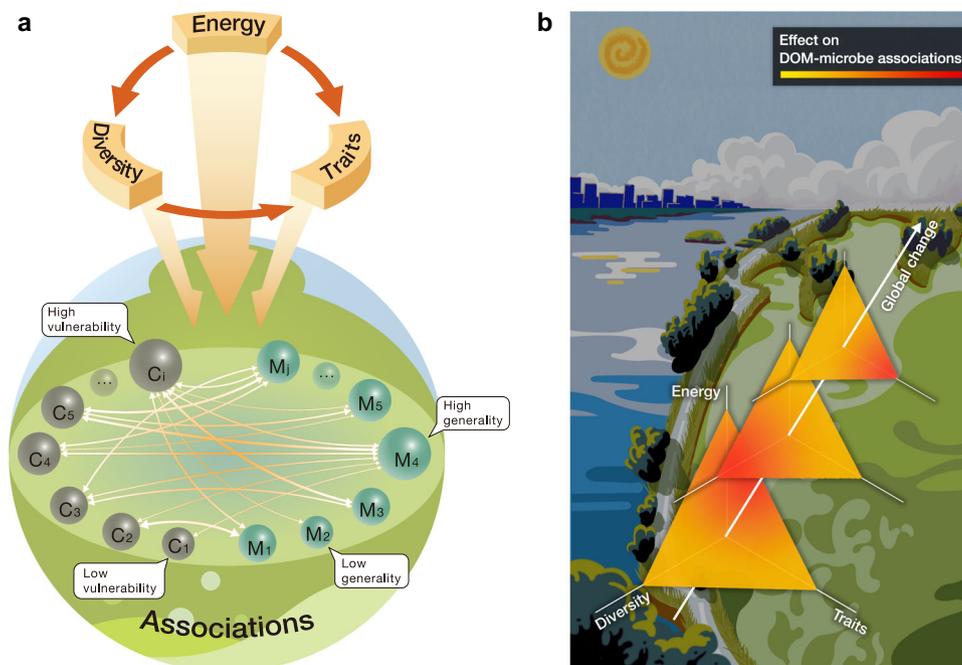
**Molecular trait:** The structural features of an individual DOM molecule deduced from molecular formula that can be assigned precisely from ultrahigh resolution mass spectrometry. These molecular-level traits include molecular weight, stoichiometry, chemical structure, oxidation state and bioavailability (e.g., lability and recalcitrance), and can be upscaled to an entire sample (i.e., compositional-level) with weighted means.

**DOM compound classes:** The groupings of DOM molecules that are categorised based on molecular traits such as H/C and O/C ratios: lipids, proteins, amino sugars, carbohydrates, unsaturated hydrocarbons, lignin, tannin, and condensed aromatics.

**DOM-microbe associations:** The relationships between DOM and microbes at multiple levels, including the compositional-level and species/molecular-level. Compositional-level associations involve correlations of alpha or beta diversity between DOM and microbes. Species/molecular-level associations refer to interactions between individual DOM molecules and microbial taxa, which can be upscaled to an interaction network by network properties, such as the measures of specialization.

**DOM-microbe bipartite network:** The interaction network between DOM and microbes, where individual organic molecules are only connected to microbial taxa that may consume or produce that specific molecule, while the direct interactions within molecules or taxa are not explicitly considered<sup>27</sup>. The network is a species/molecular-level means to quantify DOM-microbe associations and consists of either positive or negative interaction networks that are inferred by the positive and negative coefficients of correlation analyses such as SparCC<sup>35</sup>.

**DOM-microbe specialization:** A network property describing how much DOM molecules and microbial taxa interact with each other in a bipartite ecological network, that is the levels of “vulnerability” of DOM molecules and “generality” of microbial taxa. Specialization could be quantified by the standardised Shannon index  $H_2'$ <sup>28</sup> that considers the interaction frequency and strength between the two network parties. Elevated  $H_2'$  values indicate a high degree of specialization, while lower values suggest increased generalization, that is, higher vulnerability of DOM and/or higher generality of microbes (Fig. 1a).



**Fig. 1 A framework for studying the effects of global change on DOM-microbe associations.** **a** DOM-microbe associations are affected by the three proximal drivers, namely energy supply and both the diversity and traits of DOM and microbes. The relationships among the three drivers and their influences on the associations are shown with single-sided arrows. The DOM-microbe associations, indicated by double-sided arrows, are measured by bipartite interactions between DOM molecules (circles C<sub>1</sub>-C<sub>5</sub>) and microbial taxa (circles M<sub>1</sub>-M<sub>5</sub>). The size of circles indicates the abundance of DOM molecules or microbial species, and the width of arrows is the magnitude of associations. Commonly used indices summarise the specialization of individual molecule *i* and microbial species *j*, which describes the levels of “vulnerability” of DOM molecules and “generality” of microbial species. **b** Conceptual framework for understanding DOM-microbe associations under distal drivers such as global change via the three proximal drivers. For better 3D visualization, the sizes of triangles decrease towards the top-right, and the colour changes towards different corners of the triangles represent variations in the relative importance of different proximal drivers under a global change scenario. The background depicts the primary motivation of this study in examining distal drivers of climate change and eutrophication in Taihu Lake, China. The left and right waters indicate clean and cyanobacteria-dominated lake states, respectively, and are separated by a road having the shapes of western lakeshore and northern Zhushan and Meiliang Bays of Taihu Lake. We established field microcosms on mountainsides by adding sediments collected from the lake centre, and designed nutrient levels and N/P ratio based on nutrient conditions of this lake<sup>29</sup>.

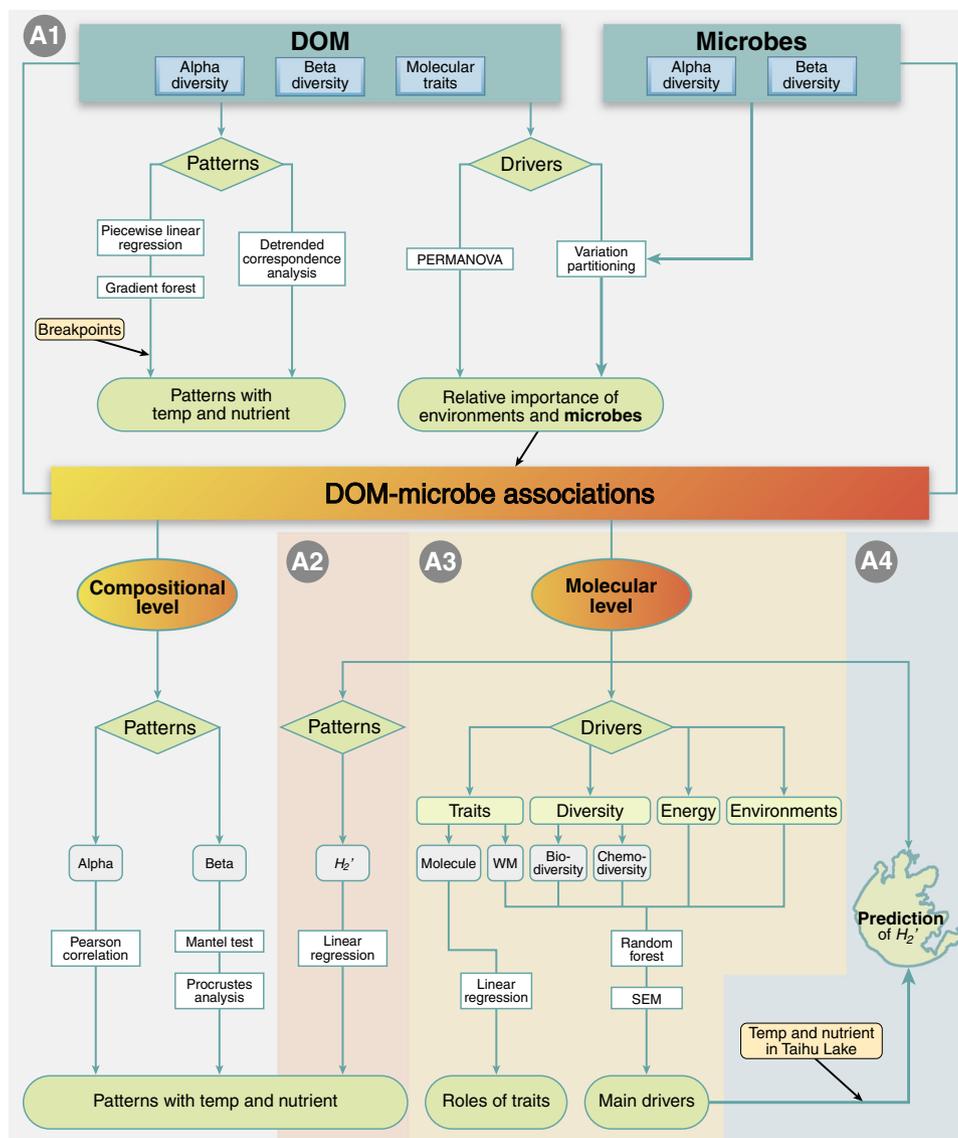
DOM. Such field experiments along natural climate gradients can disentangle temperature effects on DOM-microbe associations from nutrient enrichment, unlike in natural ecosystems where these drivers are typically confounded. The framework explicitly quantifies the relative importance of the direct effects of temperature and nutrient enrichment and the indirect effects via the proximal drivers of energy, diversity, and traits. Briefly, we selected five different elevations on each mountainside, and at each elevation, we established 30 microcosms composed of natural lake sediments and artificial lake water with ten different nutrient levels. The nutrient levels were selected based on conditions of Taihu Lake, a large eutrophic shallow lake in China with an area of 2338 km<sup>2</sup>. The sediments originated from Taihu Lake and were added to each microcosm after sterilisation to ensure identical initial DOM supply and composition. Microcosms were incubated in the field for one month allowing airborne bacteria to colonise, and sediment bacteria were examined using high-throughput sequencing of 16S rRNA genes<sup>29</sup>. In addition, we applied ultrahigh-resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to examine sediment DOM features, such as the diversity of molecular formulas (hereafter chemodiversity) and molecular traits (Table S1).

With this experiment, we address three main questions: (1) at a compositional-level, how does DOM molecular composition and its association with microbial biodiversity vary along temperature and nutrient gradients?, (2) at a molecular-level, how does the degree of specialization between DOM and microbes estimated

from interaction networks vary along temperature and nutrient gradients? and (3) how is DOM-microbe specialization inter-actively and independently influenced by temperature and nutrient enrichment via the three proximal drivers of energy, diversity, and traits? We then discuss how the effects of these drivers could be used to predict the spatiotemporal changes in specialization elsewhere under different temperature and nutrient scenarios. We focus on Taihu Lake as case study given its socioeconomic and cultural importance with over 60 million people living in the Taihu Basin, and because it is where our experimental sediments and nutrient reference conditions originated from. Together, our study advances biogeochemical modelling and improves predictions about both carbon turnover and resource-based feedbacks on microbial diversity.

## Results and discussion

**DOM features and their microbial associations at a compositional-level.** The diversity and molecular traits of DOM were strongly controlled by nutrient enrichment, and to a lesser by temperature (that is, elevation), on both subtropical and subarctic mountainsides (Figs. S1, S2). Nutrient enrichment generally increased alpha diversity (i.e., molecular richness) of DOM across all elevations when all molecular formulae were considered (Figs. 3a, S3). We identified abrupt changes in molecular composition along the nutrient gradient that mostly occurred between 1.80 and 4.05 mg N L<sup>-1</sup> for all molecules at

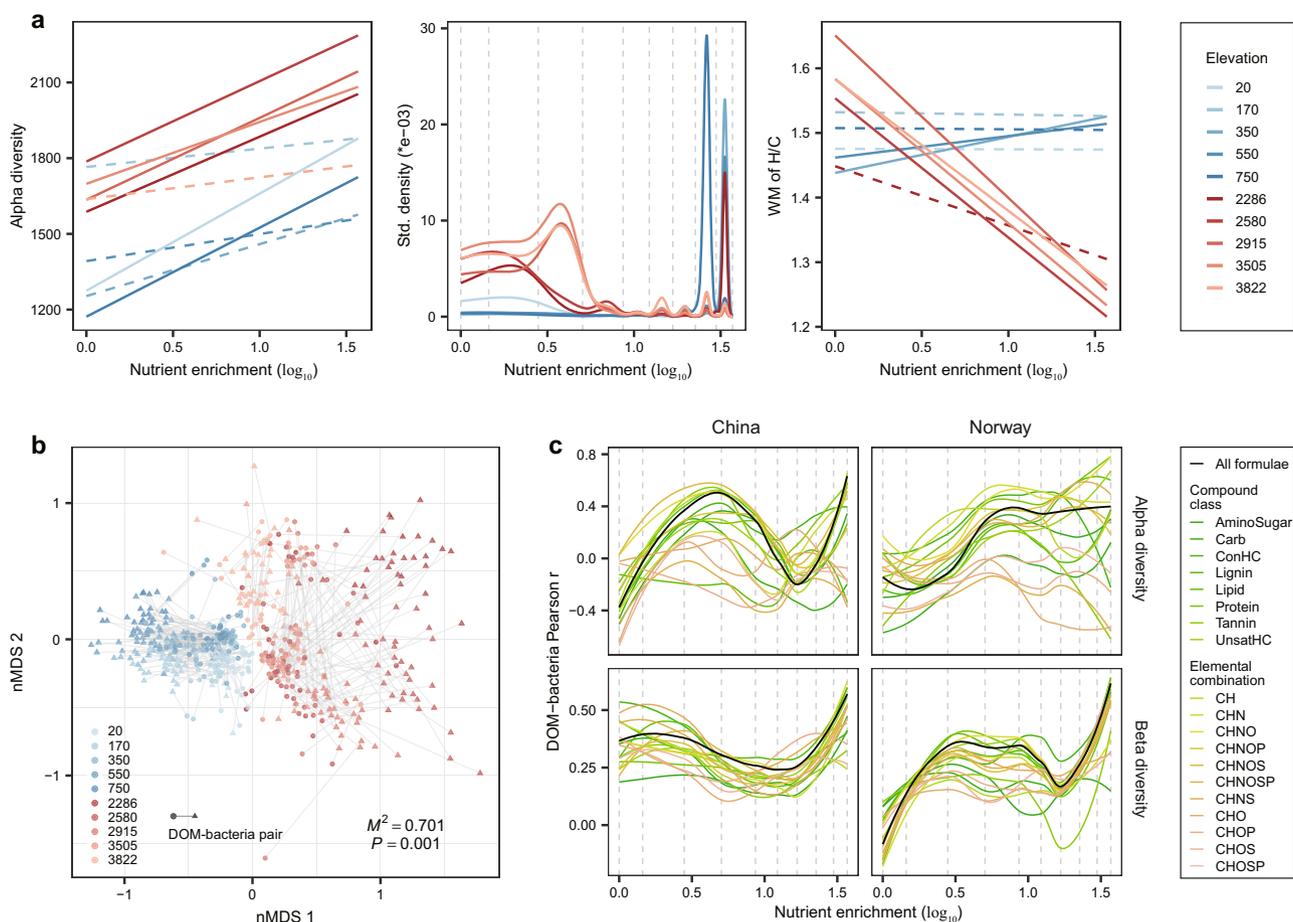


**Fig. 2 Outline of statistical analyses to address the study aims.** A1–A3 refer to the three main questions stated in the ‘Introduction’, and A4 refers to the prediction of  $H_2'$  in Taihu Lake. The response or explanatory variables in grey or light blue boxes were examined by the statistical methods in white boxes for the main purposes indicated in light green boxes or an irregular shape box. The irregular shape shows the lakeshore of Taihu Lake. PERMANOVA: permutational multivariate analysis of variance. SEM: structural equation model. Alpha: alpha-diversity of DOM and microbes. Beta: beta-diversity of either DOM or microbes. WM: weighted means of formula-based molecular traits. Temp: temperature. Details for each statistical analysis such as main purposes were described in the ‘Methods’.

each elevation using both gradient forest analysis<sup>30</sup> (Fig. 3a) and piecewise regression<sup>31</sup> (Fig. S4). The effects of nutrient enrichment on molecular traits, however, varied between the two ecoregions (Figs. 3a, S3, S5). For instance, the weighted mean of the H/C ratio in each microcosm decreased with nutrient addition to <1.5, especially at high elevations in China, indicating less bioavailable DOM (Figs. 3a, S3c). The ratio remained consistently higher ( $\geq 1.5$ ) across all nutrient levels in Norway (Figs. 3a, S3c). Given that the initial DOM composition in our study was identical everywhere, this finding suggests that the contrasting responses reflect differences in the temperature sensitivity of decomposition and/or nutrient-limited production of DOM by colonising microbes. This inability to resolve the mechanisms underlying these patterns further highlights the need for the EDTiA framework. EDTiA can help disentangle underlying drivers by explicitly quantifying the direct and indirect effects of temperature and nutrient enrichment on the associations between

DOM and microbes via intermediate environmental variables such as energy, diversity and traits (Fig. 1).

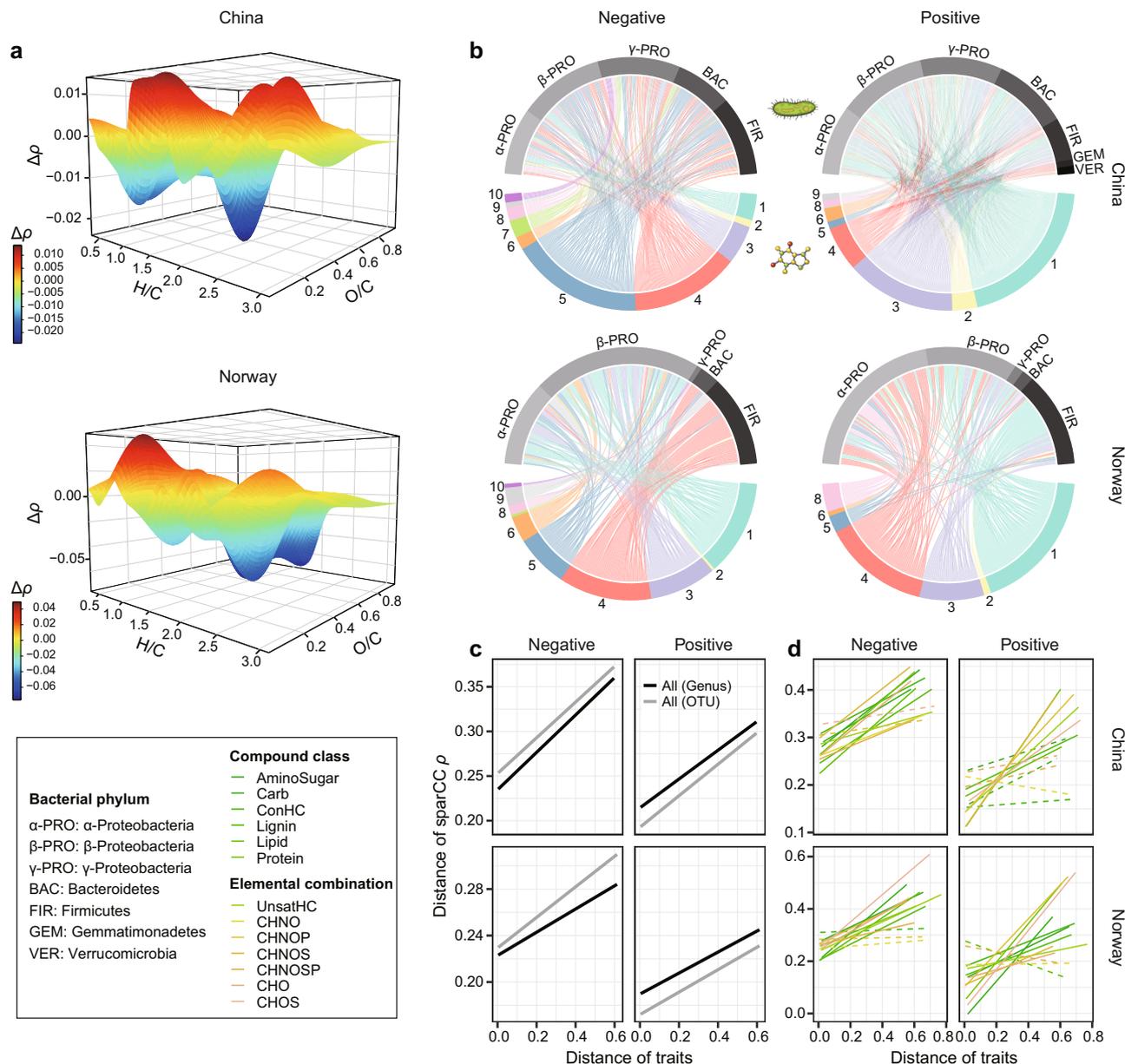
DOM composition was strongly associated with bacterial communities in both regions, and was mediated by temperature and nutrient enrichment. For instance, although environment (temperature and nutrients) and energy supply had dominant effects on DOM composition, their shared effects with biodiversity (2.7–13.1% of explained variation) indicated that these variables also indirectly influenced the associations between DOM and bacteria (Fig. S6). These DOM–bacteria associations were also supported by the fact that DOM composition was significantly predicted from bacterial community with Procrustes analysis<sup>32,33</sup> ( $M^2 = 0.701$ ,  $P \leq 0.001$ ; Fig. 3b), while their associations varied with temperature and nutrient enrichment. For example, compositional differences, indicated by the residuals of Procrustes analysis, significantly ( $P \leq 0.05$ ) decreased for all compound classes or elemental combinations at colder



**Fig. 3 DOM features and their bacterial associations at a compositional level.** **a** The effects of nutrient enrichment on DOM alpha diversity (richness), composition and molecular traits (e.g., H/C ratio) for all formulae across different elevations in China (red lines, with lower elevations in more red colours) and Norway (blue lines, with higher elevations in more blue colours). We used nitrate addition to represent nutrient enrichment as the ratio between nitrate and phosphorus in the initial overlying water was constant. Molecular richness and weighted mean (WM) of H/C ratio were plotted against the nutrient gradient of nitrate, and their relationships are indicated by solid ( $P \leq 0.05$ ) or dotted ( $P > 0.05$ ) lines estimated using linear models with one-sided F-statistics. For better visualization, we did not include the data points in Fig. 3a but showed detailed scatter plots and statistics in Fig. S3. To visualise the DOM compositional changes, we plotted the distribution of breakpoints of each molecule regarding its abundance occurring along the nutrient gradient with gradient forest analysis<sup>30</sup>. **b** The congruence between DOM and bacterial compositions across different elevations in China and Norway was examined using Procrustes analysis<sup>32,33</sup>. The NMDS plots the position of DOM and bacterial assemblages that have been optimized based on Procrustes analysis. Each line with circle and triangle ends connects to a single community of DOM and bacteria, respectively, and is coloured by elevation in either China (red) or Norway (blue). The fit of overall Procrustes transformation is reported as the  $M^2$  value, and the significance was examined using a permutation-based approach with 999 permutations and the two-sided statistically significant  $P$  value is shown. **c** The effects of nutrient enrichment on DOM-bacteria associations. The associations were quantified by the Pearson correlation coefficient  $r$  between alpha diversity of DOM and bacteria (upper panel), and by the Mantel correlation  $r$  between the beta diversity of DOM and bacteria (lower panel). We then visualised correlation  $r$  values with loess regression models along the nutrient gradient, and these correlations at each nutrient level are shown in Fig. S8. The colours of the lines indicate the DOM composition for all formulae and categories of compound classes or elemental combinations.

temperatures in China (Fig. S7). In Norway, the differences were always lower, on average, and did not vary with temperature (Fig. S7). Nutrient enrichment influenced the correlations between the number (“alpha diversity”) of DOM molecules and bacterial operational taxonomic units (OTUs) in each microcosm, or between the similarity in composition (“beta diversity”) of DOM and bacterial communities for all pairwise combinations of microcosms within a region (Fig. 3c). These results were consistent for individual compound classes or elemental combinations (Fig. 3c). Interestingly, the coordinated compositional changes in DOM and bacteria, measured by the Mantel correlation<sup>34</sup> between their beta diversities among pairwise microcosms, increased more strongly with nutrient enrichment in Norway than in China, especially at low nutrient levels beneath  $1.80 \text{ mg N L}^{-1}$  (Figs. 3c, S8).

**Ecological networks between DOM and bacteria at a molecular-level.** To quantify the associations between DOM and bacteria further from a molecular-level perspective, we first correlated the relative abundance of each DOM molecule and bacterial taxa. According to resource-consumer relationships, the negative network interactions inferred by negative correlations likely indicate the degradation of larger molecules into smaller structures, while the positive interactions may relate to the production of new molecules via degradation or biosynthetic processes. We found that the distribution of negative and positive Spearman correlations between DOM molecules and bacterial OTUs depended strongly on molecular traits. For example, molecules that were more labile, such as those with  $\text{H/C} \geq 1.5$ , were more likely to have negative correlation coefficients ( $\rho$ ) with individual OTUs ( $P \leq 0.05$ ). In contrast, molecules that were



**Fig. 4 Ecological networks between DOM and bacteria.** **a** Strength of the correlations between DOM molecules and bacterial OTUs in China (upper panel) and Norway (lower panel). For each molecule across the 150 samples in each region, we subtracted the mean absolute Spearman's rank correlation coefficient  $\rho$  of all the negative correlations with individual bacterial OTUs from the mean of the positive correlations to derive  $\Delta\rho$ .  $\Delta\rho$  was further visualised against the molecular traits H/C and O/C. **b** The negative (left panel) and positive (right panel) bipartite networks between DOM molecules and bacterial genera in China (upper panel) or Norway (lower panel) estimated using SparCC (Sparse Correlations for Compositional data)<sup>35</sup>. Upper nodes represent bacterial genera coloured by their phylum, while lower nodes represent DOM molecules coloured by the ten clusters obtained with hierarchical cluster analysis based on 16 molecular traits described in Fig. S12a and Table S1. The traits of H/C and O/C for molecules in each cluster are shown in Fig. S12b. A line connecting two nodes indicates an interaction between a DOM molecule and bacterial genus. The ten clusters are indicated by the numbers below the lower nodes. **c, d** We examined the relationships between molecular traits and the negative (left panel) or positive (right panel) DOM-bacteria bipartite networks in China (upper panel) or Norway (lower panel). Between all pairs of DOM molecules, we calculated pairwise Gower distances of their molecular traits (that is, distance of traits), or the SparCC  $\rho$  of molecules with bacterial OTUs (that is, distance of SparCC  $\rho$ ). Statistical significance between these two distance matrices was determined with a two-sided Mantel test with 999 permutations and indicated by solid ( $P \leq 0.05$ ) or dotted ( $P > 0.05$ ) lines. We considered all formulae (**c**) and also subsets of formulae within the category of compound classes or elemental combinations (**d**). For all formulae (**c**), we calculated SparCC correlation coefficients based on both bacterial OTUs (grey lines) and genera (black lines).

more recalcitrant ( $H/C < 1.5$ ) generally showed more positive correlations ( $P \leq 0.05$ ), especially in Norway (Fig. S9). These findings were even more clearly supported by the differences between the mean of the positive and negative  $\rho$  values for each molecule (Figs. 4a, S10). Correlations with individual OTUs were predominantly negative for molecules within a H/C of 1.5–2.0

and O/C of 0.4–1.0, suggesting they are largely the reactants of degradation processes, while  $\rho$  differences peaked with mainly positive values at a H/C of 1.0–1.5 and O/C of 0–0.5 indicative of in situ production (Fig. 4a).

Subsequently, we quantified DOM-bacteria interactions along temperature and nutrient gradients using the EDTiA framework.

We built bipartite networks of negative and positive interactions between DOM and bacteria at the genus level using Sparse Correlations for Compositional data (SparCC)<sup>35</sup>. SparCC can infer interaction strength by statistical correlations with a high degree of accuracy for sparse matrices, i.e. where there are many 0's<sup>35</sup>. In total, there were 6916 and 8409 interactions for negative and positive networks ( $|\text{SparCC } \rho| \geq 0.3$ ), respectively, in China, and 1313 and 2888 negative and positive interactions, respectively, in Norway (Fig. 4b). The weighted mean of the percentage of SparCC  $\rho$  values that were strongly negative ( $P \leq 0.05$ ) increased towards high nutrient levels almost exclusively in China (Fig. S11). The weighted mean of negative SparCC  $\rho$  also increased with nutrient enrichment in China (Fig. S11) and these results together suggest that nutrient enrichment increased both the number and strength of negative interactions.

We found that the negative and positive interactions strongly depended on molecular traits, which was further supported by three observations of bipartite networks. First, the negative and positive interactions were associated with different molecule groups, as categorised by a hierarchical cluster analysis of 16 molecular traits (Figs. 4b, S12). This cluster analysis identified two groups of molecules (clusters 4 and 5) that were largely recalcitrant with a H/C of  $<1.5$ , while two separate groups (clusters 1 and 3) mostly included labile molecules with a H/C of  $\geq 1.5$  (Fig. S12). In China, negative interactions were dominant between the molecules in clusters 4 or 5 and the bacteria in phyla Proteobacteria, Bacteroidetes, or Firmicutes, while the positive interactions were mostly linked to the molecule clusters 1 and 3 (Fig. 4b). In Norway, the molecule cluster 4 was mainly negatively linked to Firmicutes and positively linked to  $\alpha$ - and  $\beta$ -Proteobacteria (Fig. 4b). Second, molecules generally covaried more similarly with bacteria as they were more similar in their traits. This was indicated by the positive linear slopes between the pairwise differences in both the traits of DOM molecules and their interaction strength measured by SparCC  $\rho$  values (Mantel test,  $P \leq 0.001$  in each region; Fig. 4c). For example, had DOM molecules more similar H/C or O/C values, they correlated more similarly with bacterial taxa. Third, molecular traits were more strongly correlated with the strength of DOM-bacteria interactions in the negative than positive interaction networks for all molecules (Fig. 4c), which was also true for most of the networks when considering compound classes or elemental combinations (Fig. 4d). These correlations, consistent at both the genus and OTU levels (Fig. 4c), indicate that molecular traits may be better at predicting the decomposition than production of DOM.

Finally, we calculated the degree of specialization between DOM and bacteria in the entire negative and positive interaction networks using the  $H_2'$  index<sup>28</sup>. We also calculated specialization  $d'$  indices for individual DOM molecules and bacterial genera<sup>28</sup>. Elevated  $H_2'$  or  $d'$  values indicate a high degree of specialization, while lower values suggest increased generalization. That is, highly specialized DOM or bacteria, in the extreme example, mean that a single DOM molecule is used by one bacterial taxon, whereas highly generalized DOM or bacteria mean that different DOM molecules can be used by a large number of bacterial taxa (Fig. 1a). We found that as the negative networks between DOM and bacteria became more specialized (i.e., higher  $H_2'$  values), they corresponded with more specialized DOM molecules (i.e., higher weighted mean  $d'$ ; Fig. S13). More specialized decomposition processes can therefore reduce the vulnerability of DOM to degradation as they will require highly specialized consumers. For positive networks,  $H_2'$  values showed consistent patterns along nutrient enrichment gradients with those of  $d'$  for both DOM and bacteria (Fig. S13). This result indicates that increased specialization in production processes was related to the decreased vulnerability of DOM and the decreased generality of microbes,

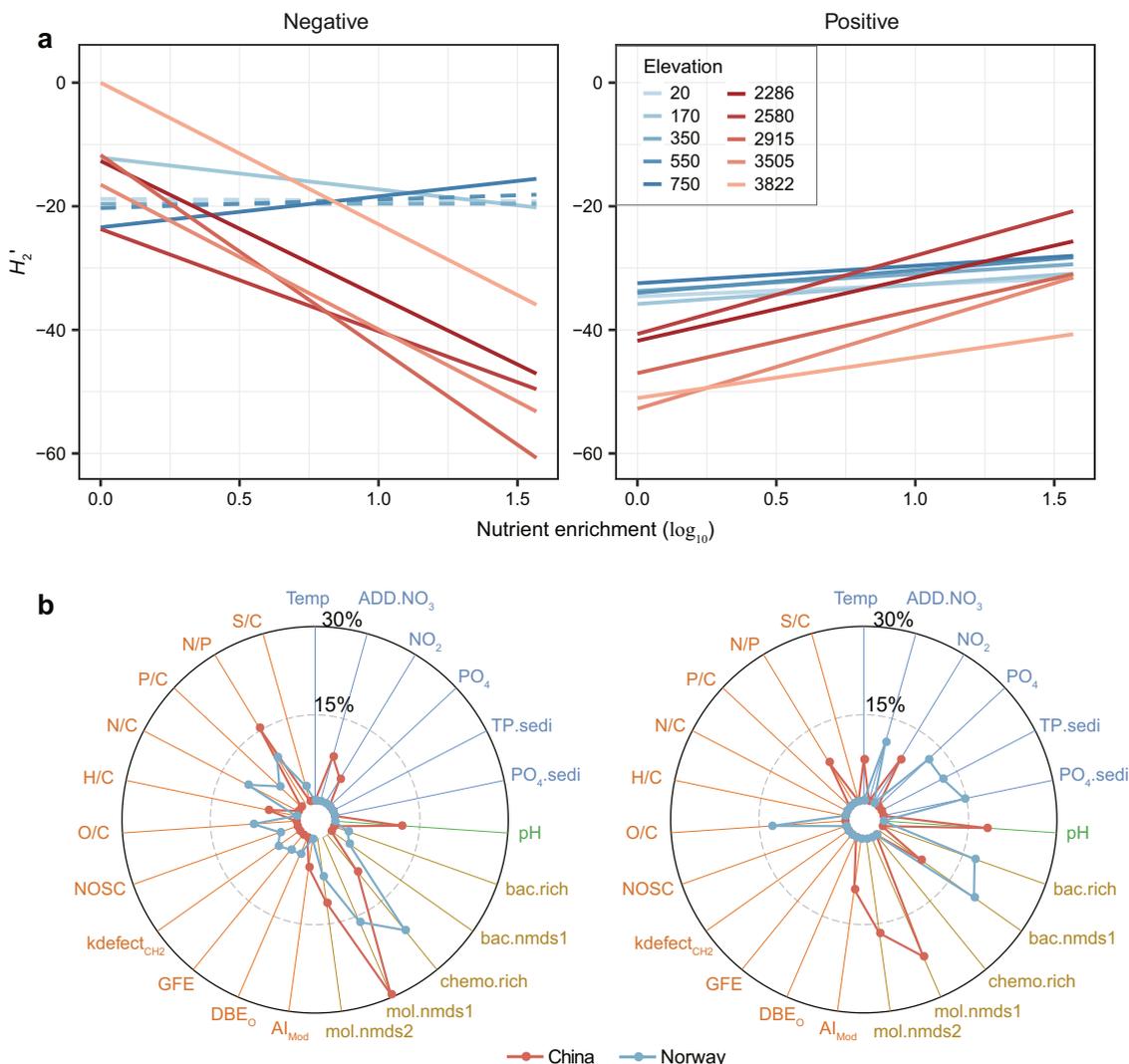
and thus the potentially decreased DOM production. These results collectively suggest that in addition to the specialization perspective of bacteria or DOM,  $H_2'$  can summarise resource-consumer relationships at an ecosystem-level. In both regions,  $H_2'$  was higher, on average, in negative than positive interaction networks ( $t$ -test,  $t = 2.11$ ,  $P = 0.04$  in China and  $t = 23.57$ ,  $P \leq 0.001$  in Norway; Figs. 5a, S14), indicating greater specialization in the decomposition than production processes of microbes. The mean specialization  $H_2'$  of negative ( $t$ -test,  $t = -10.19$ ,  $P \leq 0.001$ ) and positive ( $t$ -test,  $t = -6.56$ ,  $P \leq 0.001$ ) networks were also significantly higher in Norway than in China (Fig. 5a), suggesting more specialized decomposition (i.e., negative networks) and thus potentially lower vulnerability of DOM decomposition to bacteria in subarctic regions.

Nutrient enrichment showed divergent effects on the  $H_2'$  of negative or positive interaction networks between the two study regions. Specifically, nutrient enrichment substantially decreased the  $H_2'$  of negative networks for all molecules in China (Fig. 5a), which was particularly true when considering only recalcitrant components, such as lignin and CHNO (Fig. S15). Compared to Norway, nutrient enrichment increased the  $H_2'$  of positive interactions relatively more at lower elevations in China (Fig. 5a). Nutrient enrichment in the subtropical region could thus contribute to the greater recalcitrance of DOM by increased decomposition (i.e., less specialized negative networks) and reduced production of molecules (i.e., more specialized positive networks).

**Drivers of DOM-bacteria interaction networks.** We explored the effects of distal and proximal drivers on negative and positive DOM-bacteria networks using the EDTiA framework (Fig. 1). The distal drivers were temperature and nutrient enrichment as proxies of climate change and human impacts, respectively. The three proximal drivers were energy supply, such as primary productivity and sediment total organic carbon, the diversity of bacteria and DOM, that is the richness and composition of bacteria and DOM, and DOM molecular traits (Table S1).

We found that, in addition to bacterial diversity and chemodiversity, molecular traits also influenced the network specialization  $H_2'$ . In the negative networks,  $H_2'$  was most strongly correlated with DOM molecular composition ( $r = 0.77$ ,  $P \leq 0.001$ ), followed by molecular richness ( $r = -0.76$ ,  $P \leq 0.001$ ) and molecular N/P ratio ( $r = 0.76$ ,  $P \leq 0.001$ , Fig. S16). In contrast, in the positive networks,  $H_2'$  was less correlated with molecular traits (Fig. S16). Furthermore, we examined the relative importance of diversity and molecular traits for predicting  $H_2'$ .  $H_2'$  was mainly affected by chemodiversity, such as molecular richness or DOM composition, followed by molecular traits, such as N/P or N/C ratios, in the negative networks, whereas chemodiversity, biodiversity, environmental variables and energy supply were all similarly important in the positive networks (Fig. 5b).

We finally tested the hypothesised effects of two global change drivers, temperature change and nutrient enrichment, on the specialization of DOM-bacteria interaction networks. We compared these effects to other drivers such as contemporary nutrients, energy supply, biodiversity, chemodiversity and molecular traits using structural equation models (SEM)<sup>36</sup> (Fig. 6a). SEM tests hypothesized cause-and-effect relationships, which are translated into regression equations and fitted to data. Through this process, SEM separates the direct and indirect effects of different drivers of  $H_2'$ . The SEM results indicated that there were different constraints on the specialization between negative and positive interactions. For the negative interactions, both global change drivers strongly influenced  $H_2'$  through

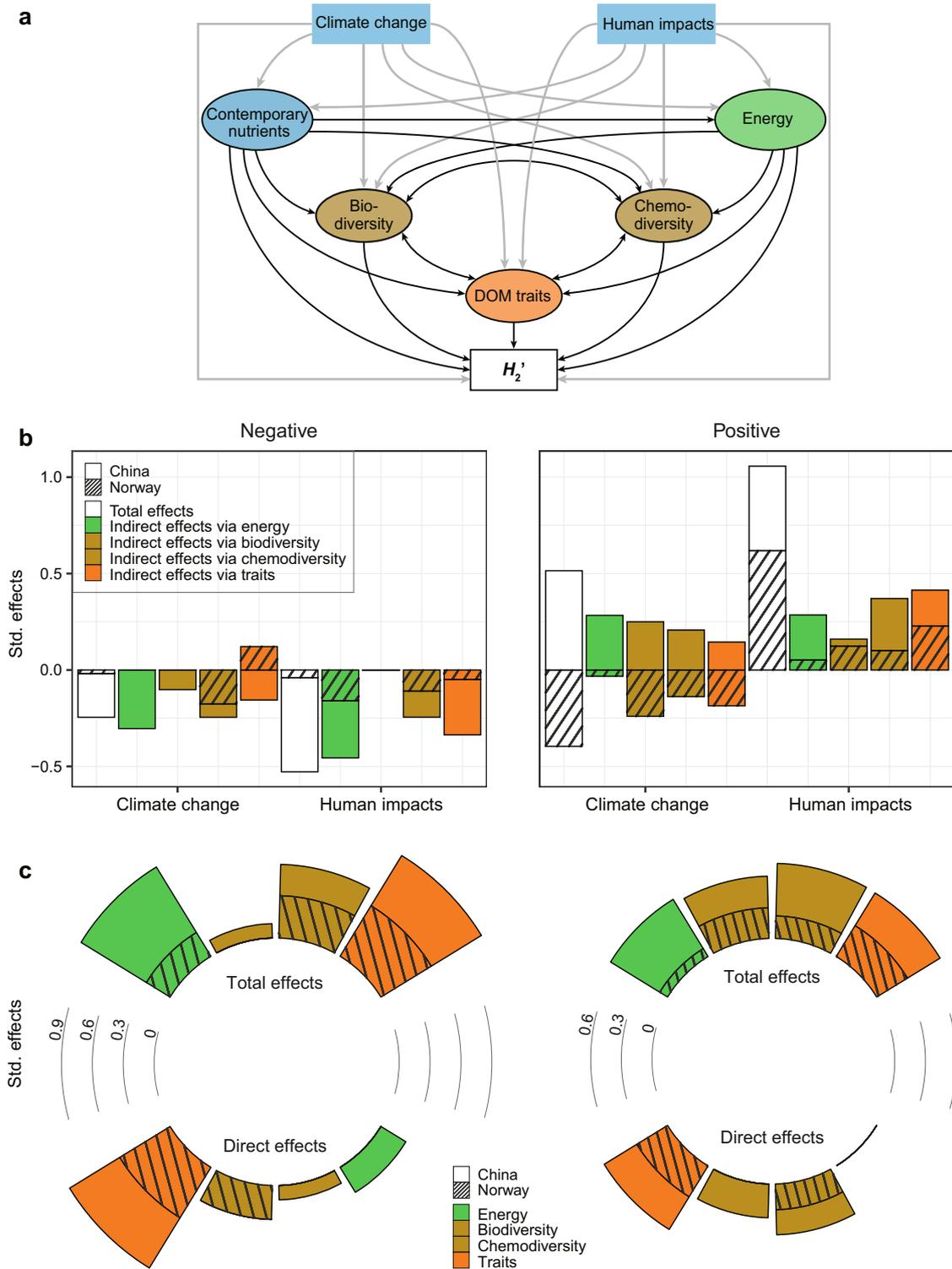


**Fig. 5** Relative importance of diversity and molecular traits in explaining specialization of DOM-bacteria networks. **a** We plotted specialization  $H_2'$  against nutrient enrichment for negative (left panel) and positive (right panel) bipartite networks for each elevation in China (red lines) and Norway (blue lines). Statistical significance of linear model fits with one-sided F-statistics was indicated by solid ( $P \leq 0.05$ ) or dotted ( $P > 0.05$ ) lines. For better visualization, we omitted the data points but detailed scatter plots and statistics are shown in Fig. S14. **b** We examined the relative importance of all explanatory variables on the  $H_2'$  of negative (left panel) and positive (right panel) bipartite networks in China (red lines) and Norway (blue lines) using random forest. The relative contribution (%) of each variable towards  $H_2'$  is shown in radar plots. The explanatory variables were grouped by environment, energy, diversity and traits with consistent colours of ovals or rectangles as in Fig. 6a. Abbreviations of explanatory variables are detailed in Table S1.

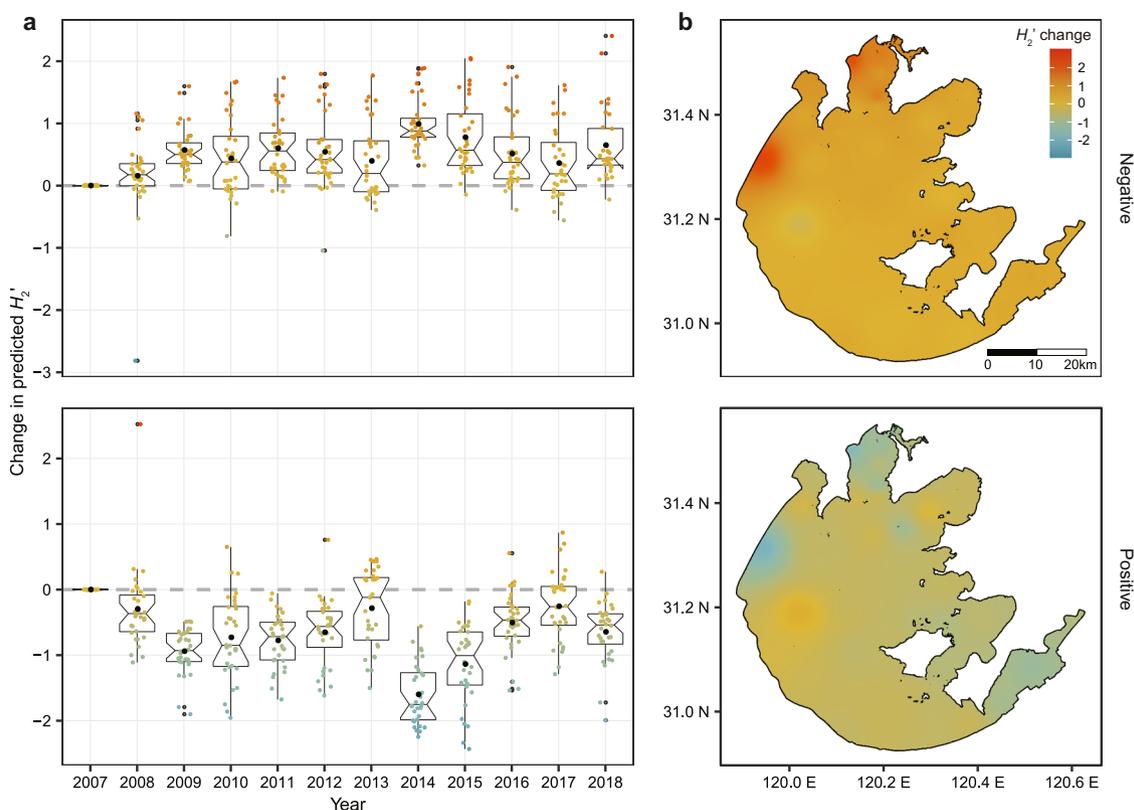
indirect effects on energy supply and molecular traits, especially in China (Figs. 6b, c, S17). In contrast to Norway, both climate change and human impacts had larger total mean effects of  $-0.23$  and  $-0.49$ , respectively, on the  $H_2'$  of negative interactions in China (Fig. 6a). However, molecular traits had the dominant direct effects on  $H_2'$  in both China and Norway, with similar mean standardised effect size of  $0.57$  ( $P \leq 0.001$ ; Figs. 6b, S17). For the positive interactions, there were large total mean effects of climate change ( $0.51$  and  $-0.40$  for China and Norway, respectively) and human impacts ( $0.44$  and  $0.62$ , respectively), both of which indirectly influenced  $H_2'$  similarly through biodiversity, chemodiversity and molecular traits (Figs. 6b, c, S17).

**Prediction of DOM-bacteria interactions.** As inland waters worldwide are affected by climate change<sup>37</sup> and undergo changes in trophic status<sup>38</sup>, our approaches could be used to make predictions about how microbes degrade and produce DOM. For

instance, total nitrogen in Taihu Lake has been reduced by a mean ( $\pm$ SD) of  $1.24$  ( $\pm 1.41$ )  $\text{mg L}^{-1}$  via lake management efforts since a hyper-eutrophication event occurred in May 2007 (Fig. S18). Based on the estimated direct and indirect effects of distal drivers in the SEM fitted to the Chinese data (Fig. 6a), this oligotrophication, combined with a mean decrease in water temperature of  $0.20$  ( $\pm 0.87$ )  $^{\circ}\text{C}$  between 2007 and 2018, was predicted to shift DOM-bacteria interaction networks towards more specialized DOM decomposition and more generalized DOM production (Fig. 7). This result indicates that lake management has potentially increased carbon stocks in the lake. Specifically,  $H_2'$  changed by  $+0.65$  ( $\pm 0.58$ ) and  $-0.65$  ( $\pm 0.46$ ) for negative and positive interactions, respectively, over this period (Fig. 7a). The greatest changes happened in the most eutrophic part of the lake, including the northwestern lakeshore and the northern Zhushan and Meiliang Bays (Figs. 7b, S19). Although our predictions lack detailed spatiotemporal environmental variation, as used to parameterize the SEM models, they do illustrate the potential to upscale our findings in real-world settings. Ground-truthing our



**Fig. 6 Structural equation models to explain specialization of DOM-bacteria networks.** **a** Conceptual model showing the hypothesized causal relationships among distal and proximal drivers for the specialization  $H_2'$  of DOM-bacteria interaction networks. The distal drivers include climate change, human impacts and contemporary nutrients, and the proximal drivers are energy supply, biodiversity, chemodiversity and molecular traits, which are described in detail in Table S3. Grey or black arrows indicate the hypothesized relationships among the exogenous or endogenous composite variables and  $H_2'$ , respectively. **b, c** Stacked bar plots show the standardised effects (Std. effects) of predictor variables on the  $H_2'$  of negative (left panel) and positive (right panel) bipartite networks in China or Norway estimated from the best supported models. We considered **b** the total and indirect effects of global change and human impacts via proximal variables and **c** the total and direct effects of proximal variables. Details of the full structural equation models are shown in Fig. S17.



**Fig. 7** Decadal change in predicted specialization of DOM-bacteria networks in Taihu Lake. **a** Changes in  $H_2'$  of negative (upper panel) and positive (lower panel) bipartite networks from 2007 to 2018. **b** The spatial distribution of changes in  $H_2'$  of negative (upper panel) and positive (lower panel) networks in 2018 across the Taihu Lake. Estimated changes in  $H_2'$  were calculated ( $n = 32$  sampling sites across the whole of Taihu Lake; Fig. S19a) by comparing with the baseline of 2007, and represent the combined effects of climate change and eutrophication. The colored dots in **a** indicate  $H_2'$  changes for individual sites which are consistent with the figure legend of **(b)**, and black dots are the mean values for each year. The box in **a** represents the interquartile (50% of data), the horizontal line in the box represents the median, the “notch” represents the 95% confidence interval of the median and the “whiskers” represent the maximum and minimum values.

results with in situ measurements across environmental gradients and across spatiotemporal scales could further validate our predictions. Our predictions also highlight how lake management and policy can affect the balance between decomposition and production of organic matter and lake carbon cycling, more generally. Future studies with more lakes or other aquatic systems are needed for a comprehensive understanding of how global change will shift DOM-bacteria interaction networks in inland waters worldwide.

**Implications.** The factors that control microbial processing of DOM composition, and consequently its degradation, are complex and challenging to discern<sup>39</sup>, yet are critical for predicting carbon cycling under global change scenarios. We found that associations between DOM and microbial decomposers were influenced by fundamental drivers of ecosystem functioning, such as energy supply<sup>19,20</sup>, both DOM and microbial compositions<sup>9,14</sup> and molecular traits<sup>8</sup>. The EDTiA framework we developed provides a unified approach from a molecular-level perspective to identify when each of these proximal drivers is more important, and to separate contrasting biological processes associated with DOM degradation and production that may have obscured previous analyses of bulk DOM pools. In addition to energy supply and the diversity of DOM and bacteria, our study reveals that molecular traits are informative for describing DOM-bacteria interaction networks across contrasting climatic zones, especially the negative interactions indicative of degradation processes. Although molecular traits are well known to be linked with DOM

persistence or vulnerability to degradation<sup>8,40</sup>, their influence on the underlying biological mechanisms has remained poorly understood. Our results advance this work by demonstrating when the specialization of DOM-microbe networks changes with molecular traits, and by providing predictions of how specialization might vary under global change scenarios.

We found that temperature and nutrient enrichment can change DOM-bacteria interactions by shifting the three proximal drivers, namely energy, diversity, and traits. For the positive bipartite networks, nutrient enrichment generally increased the DOM-bacteria specialization, and more so than temperature, by changing biodiversity, chemodiversity, and molecular traits. Positive interactions related to the production of new molecules depend on the specific interacting partners, which is partly supported by the positive relationships between  $H_2'$  and the  $d'$  of DOM or bacteria (Fig. S13). In the negative networks, however, both temperature and nutrient enrichment reduced specialization, primarily via changing molecular traits and energy supply. The greater importance of molecular traits indicates that decomposition processes associated with negative networks may depend more on whether molecules contain structures that resist degradation<sup>8</sup>, especially in the presence of temperature and nutrient limitation<sup>41</sup>. At lower temperatures and nutrient levels, the required activation energy to degrade these molecules may become more limiting<sup>41</sup>. We also found that the importance of these distal drivers of climate change and human impacts varied between biomes. For instance, both elevated temperature and nutrient enrichment reduced the specialization of negative

DOM-bacteria interactions in subtropical China, but these two drivers were less important in subarctic Norway. As their indirect effects via microbial composition varied between biomes, these responses may partly reflect differences in the functional traits and activity of biological communities. Such mechanisms via the three proximal drivers may also help evaluate how the biological approaches such as biochemical oxygen demand could be applied in assessing dissolved organic carbon dynamics in natural waters<sup>42</sup>. Future studies with metagenomics and metatranscriptomics could offer a powerful complement to test how microbial traits vary with DOM traits.

In summary, we found that DOM-microbe associations were primarily influenced by temperature change and nutrient enrichment via energy, diversity and traits, the integration of which is a requisite for predicting how organic carbon responds to multiple global change drivers. Looking forward, there is a need to translate the DOM-microbe associations of our EDTiA framework into process-based ecosystem models, from which predictions of the future carbon cycle stand to be improved by incorporating more information on microbial community function, such as their specialization on different DOM sources<sup>43,44</sup>. More generally, our work shows how the molecular traits of DOM control the responses of DOM-microbe networks and their associated biogeochemical cycles in a changing world.

## Methods

**Experimental design.** The comparative field microcosm experiments were conducted on Laojun Mountain in China (26.6959 N; 99.7759 E) in September–October 2013, and on Balggesvarri Mountain in Norway (69.3809 N; 20.3483 E) in July 2013, designed to be broadly representative of subtropical and subarctic climatic zones, respectively, as first reported in Wang et al.<sup>29</sup>. In the Laojun Mountain region, mean annual temperatures ranged from 4.2 to 12.9 °C, with July mean temperatures of 17–25 °C. In the Balggesvarri Mountain region, mean annual temperatures ranged from –2.9 to 0.7 °C, with July mean temperatures of 8–16 °C. The experiments were characterised by an aquatic ecosystem with consistent initial DOM composition but different locally colonised microbial communities and newly produced endogenous DOM. While allowing us to minimise the complexity of natural ecosystems, the experiment provided a means for investigating DOM-microbe associations at large spatial scales by controlling the initial DOM supply. Briefly, we selected locations with five different elevations on each mountainside. The elevations were 3822, 3505, 2915, 2580 and 2286 m a.s.l. on Laojun Mountain in China, and 750, 550, 350, 170 and 20 m a.s.l. on Balggesvarri Mountain in Norway. At each elevation, we established 30 aquatic microcosms (1.5 L bottle) composed of 15 g of sterilised lake sediment and 1.2 L of sterilised artificial lake water at one of ten nutrient levels of 0, 0.45, 1.80, 4.05, 7.65, 11.25, 15.75, 21.60, 28.80 and 36.00 mg N L<sup>-1</sup> of KNO<sub>3</sub> in the overlying water. To compensate for nitrate additions shifting stoichiometric ratios, KH<sub>2</sub>PO<sub>4</sub> was added to the bottles so that the N/P ratio of the initial overlying water was 14.93, which was similar to the annual average ratio in Taihu Lake during 2007 (that is, 14.49). Thus, we use “nutrient enrichment” to indicate a series of targeted nutrient levels of both nitrate and phosphate, the former of which was used to represent nutrient enrichment in the statistical analyses. Each nutrient level was replicated three times. The lake sediments were obtained from the centre of Taihu Lake, China, and were aseptically canned per bottle after autoclaving at 121 °C for 30 min. Nutrient levels for the experiments were selected based on conditions of the eutrophic Taihu Lake, and the highest nitrate concentration was based on the maximum total nitrogen in 2007 (20.79 mg L<sup>-1</sup>; Fig. S19). We chose the nutrient level of this year because a massive cyanobacteria bloom in Taihu Lake happened in May 2007 and initiated an odorous drinking water crisis in the nearby city of Wuxi.

The microcosms were left in the field for one month allowing airborne bacteria to freely colonise the sediments and water. To keep the microbial dispersal events as natural as possible, we did not cover the experimental microcosms in case of rainfall. To avoid or minimize potential influence of extreme nature events, we (i) left the top 20% of each microcosm empty to prevent water from overflowing during heavy rains, and (ii) checked the experimental sites twice during each experimental period, and added sterilized water to obtain a final volume of approximately 1.2 L. The bottom of our microcosm was buried into the local soils by 10% of the bottle height, partly to reduce UV exposure to sediments. More considerations of the experimental design were detailed in the Supplementary Methods. To avoid the effects of daily temperature variation, we measured the water temperature and pH within 2 h before noon at all elevations in the day before the final sample collection. At the end of the experimental period, we aseptically sampled the water and sediments of the 300 bottles (that is, 2 mountains × 5 elevations × 10 nutrient levels × 3 replicates) for the following analyses of physicochemical variables, bacterial community and DOM composition.

**Physicochemical variables and bacterial community.** We measured environmental variables, namely, the total nitrogen (TN), total phosphorus (TP), dissolved nutrients (that is, NO<sub>x</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup> and PO<sub>4</sub><sup>3-</sup>), total organic carbon (TOC), dissolved organic carbon (DOC) and chlorophyll *a* (Chl *a*) in the sediments, and the NO<sub>3</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, PO<sub>4</sub><sup>3-</sup> and pH in the overlying water (Table S2, Fig. S20), according to Wang et al.<sup>29</sup>.

The sediment bacteria were examined using high-throughput sequencing of 16S rRNA genes. The sequences were processed in QIIME (v1.9)<sup>45</sup> and OTUs were defined at 97% sequence similarity. The bacterial sequences were rarefied to 20,000 per sample. Further details on physicochemical and bacterial community analyses are available in Wang et al.<sup>29</sup>.

**ESI FT-ICR MS analysis of DOM samples.** Highly accurate mass measurements of DOM within the sediment samples were conducted using a 15 Tesla solarix XR system, a ultrahigh-resolution Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS, Bruker Daltonics, Billerica, MA) coupled with an electrospray ionization (ESI) interface, as demonstrated previously<sup>46</sup> with some modifications. It should be noted that FT-ICR MS does not identify molecules, but only molecular formulae in terms of elemental composition and there can be many molecular structures sharing the same elemental compositions. DOM was solid-phase extracted (SPE) with Agilent VacElut resins before FT-ICR MS measurement<sup>47</sup> with minor modifications. Briefly, an aliquot of 0.7 g freeze-dried sediment was sonicated with 30 ml ultrapure water for 2 h, and centrifuged at 5000 × *g* for 20 min. The extracted water was filtered through the 0.45 μm Millipore filter and further acidified to pH 2 using 1 M HCl. Cartridges were drained, rinsed with ultrapure water and methanol (ULC-MS grade), and conditioned with pH 2 ultrapure water. Calculated volumes of extracts were slowly passed through cartridges based on DOC concentration. Cartridges were rinsed with pH 2 ultrapure water and dried with N<sub>2</sub> gas. Samples were finally eluted with methanol into precombusted amber glass vials, dried with N<sub>2</sub> gas and stored at –20 °C until DOM analysis. The extracts were continuously injected into the standard ESI source with a flow rate of 2 μl min<sup>-1</sup> and an ESI capillary voltage of 3.5 kV in negative ion mode. One hundred single scans with a transient size of 4 mega word (MW) data points, an ion accumulation time of 0.3 s, and within the mass range of *m/z* 150–1200, were co-added to a spectrum with absorption mode for phase correction, thereby resulting in a resolving power of 750,000 (FWHM at *m/z* 400). All FT-ICR mass spectra were internally calibrated using organic matter homologous series separated by 14 Da (–CH<sub>2</sub> groups). The mass measurement accuracy was typically within 1 ppm for singly charged ions across a broad *m/z* range (150–1200 *m/z*).

Data Analysis software (BrukerDaltonik v4.2) was used to convert raw spectra to a list of *m/z* values using FT-MS peak picker with a signal-to-noise ratio (S/N) threshold set to 7 and absolute intensity threshold to the default value of 100. Putative chemical formulae were assigned using the software Formularity (v1.0)<sup>48</sup> following the Compound Identification Algorithm<sup>49</sup>. In total, 19,538 molecular formulae were putatively assigned for all samples (*n* = 300) based on the following criteria: S/N > 7, and mass measurement error < 1 ppm, considering the presence of C, H, O, N, S and P and excluding other elements or an isotopic signature. All formula assignments were further screened to meet the criteria as follows<sup>50</sup>: (1) formulae containing an odd number of nitrogen atoms had an even nominal *m/z* and those containing an even number of nitrogen atoms had an odd nominal *m/z*; (2) the number of hydrogen atoms was at least 1/3 of carbon and could not exceed 2C + N + 2; (3) the number of nitrogen or oxygen atoms could not exceed the number of carbon atoms; (4) the ratio of O/C was set to 0–1, H/C ≥ 0.3, N/C ≤ 1, double bond equivalents (DBE) ≥ 0.

The assigned molecules were categorised into eight compound classes or 12 elemental combinations. The compound classes based on van Krevelen diagrams<sup>51</sup> were lipids (O/C = 0–0.3, H/C = 1.5–2.0), proteins (O/C = 0.3–0.55, H/C = 1.5–2.2), amino sugars (O/C = 0.55–0.67, H/C = 1.5–2.2), carbohydrates (Carb; O/C = 0.67–1.2, H/C = 1.5–2), unsaturated hydrocarbons (UnsatHC; O/C = 0–0.1, H/C = 0.7–1.5), lignin (O/C = 0.1–0.67, H/C = 0.7–1.5), tannin (O/C = 0.67–1.2, H/C = 0.5–1.5) and condensed aromatics (ConHC; O/C = 0–0.67, H/C = 0.2–0.7). The elemental combinations were CH, CHN, CHNO, CHNOP, CHNOS, CHNOSP, CHNS, CHO, CHOP, CHOS, CHOSP and CHS.

**Estimating DOM features.** We considered DOM features from three aspects: alpha diversity, beta diversity and molecular traits. These features were considered for all molecules (19,538 different formulae), but also for subsets of molecules within each category of compound classes or elemental combinations. The dataset based on all molecular formulae was named “All molecules”, while the datasets of subsets of formulae were named by “category name + compounds”. The relative abundance of molecules was calculated by normalizing signal intensities of assigned peaks to the sum of all intensities within each sample. Peak (i.e., molecule) intensity may not always necessarily provide reliable information on absolute concentrations. Therefore, we did not analyse absolute intensities but instead standardised the intensity relative to all other molecules in a sample. The measure of relative abundance has been shown to be very informative in revealing ecological patterns in previous studies, such as along soil depth<sup>9</sup> and in lakes<sup>8</sup>, and so should help determine how temperature and nutrient enrichment influence DOM-microbe associations. Notably, there are also methodological reports supporting the

quantitative aspect of high-resolution mass spectrometry method. For instance, the intensity of compounds varies linearly with their initial concentration, and this linear relationship was quite similar among compounds, especially when coupled with charged aerosol detection<sup>52</sup>. More importantly, our experimental design used a common pool of homogenized sediments across experimental units such that the overall matrix of organic molecules may be assumed to be relatively similar across samples. This provides a greater confidence that a change in peak intensity for a given molecule reflects a relative change in its concentration across microcosms (i.e., samples) and the two main environmental gradients.

We considered two aspects of chemodiversity: chemical alpha diversity and chemical beta diversity. Chemical alpha diversity was calculated using a richness index that counts the total number of peaks in each sample. Chemical beta diversity was calculated with the Bray-Curtis dissimilarity metric, and further represented by the first two axes of a non-metric multidimensional scaling (NMDS) ordination of this dissimilarity. We also considered overall molecular composition, which was visualised across the elevations and nutrient enrichment treatments with detrended correspondence analysis (DCA)<sup>53</sup>. The analyses of chemical diversity were performed using the R package `vegan` V2.4.6<sup>54</sup>.

We also calculated 16 molecular traits that could affect microbial associations and were related to molecular weight, stoichiometry, chemical structure, and oxidation state (Table S1). These traits were mass, the number of carbon (C) atoms, the modified aromaticity index ( $AI_{Mod}$ )<sup>55</sup>, DBE<sup>55</sup>, DBE minus oxygen ( $DBE_O$ )<sup>55</sup>, DBE minus AI ( $DBE_{AI}$ )<sup>55</sup>, standard Gibb's Free Energy of carbon oxidation (GFE)<sup>56</sup>, Kendrick Defect ( $kdefect_{CH_2}$ )<sup>57</sup>, nominal oxidation state of carbon (NOSC), O/C ratio, H/C ratio, N/C ratio, P/C ratio, S/C ratio, and carbon use efficiency ( $Y_{met}$ )<sup>58</sup>. All calculations were performed using the R package `fmsRanalysis` V1.0.0<sup>59</sup> and the scripts at <https://github.com/danczakre/ICRTutorial>. DBE represents the number of unsaturated bonds and rings in a molecule<sup>55</sup>. Higher values of DBE, AI and NOSC all indicate a higher recalcitrance of DOM. A large Kendrick Defect can indicate a higher degree of oxidation. Lower values of  $Y_{met}$  indicate a higher thermodynamic efficiency of metabolic reactions involved in biomass production<sup>58</sup>. Weighted means of formula-based molecular traits (for example the  $Mass_{wm}$  for Mass) were calculated as the sum of the product of the trait value for each individual molecule ( $Mass_i$ ) and relative intensity  $I_i$  divided by the sum of all intensities with the R package `FD` V1.0.12<sup>60</sup> using the equation:

$$Mass_{wm} = \frac{\sum Mass_i \times I_i}{\sum I_i} \quad (1)$$

In addition, ten clusters of molecules were grouped based on the 16 molecular traits by hierarchical cluster analysis using Ward's minimum variance method with the R package `stats` V3.6.1.

**Characterising bacterial communities.** The relative abundance of OTUs was calculated by the normalization of read counts of OTUs to the sum of all reads within each sample. Likewise, we considered two aspects of biodiversity: bacterial alpha diversity and beta diversity. Bacterial alpha diversity was calculated using species richness that counts the total number of OTUs in each sample. Bacterial beta diversity was calculated with the Bray-Curtis dissimilarity metric, and further represented by the first two axes of NMDS of this dissimilarity.

**Estimating associations between DOM and bacteria.** At the DOM compositional level, we examined DOM-bacteria associations from the following aspects: Pearson's correlation between alpha diversity of DOM and bacteria, and a Mantel correlation between the beta diversity of DOM and bacteria (Box 1). We also tested the congruence between DOM and bacterial composition using Procrustes analysis of NMDS coordinates estimated for each community across elevations and nutrient enrichment levels with the Bray-Curtis dissimilarity metric<sup>32,33</sup>. Procrustes analysis is a technique for comparing the relative positions of points (i.e., samples or sites) in two multivariate datasets (in an ordination space). It attempts to stretch and rotate the points in one matrix, such as points obtained from a NMDS, to be as close as possible to points in another matrix, thus preserving the relative distances between points within each matrix<sup>32,33</sup>. This procedure yields a measure of fit,  $M^2$ , which is the sum of squared distances between corresponding data points after the transformation. Analogous to a Mantel test, Procrustes analysis is particularly used to determine how much variance in one matrix (i.e., bacteria) is attributable to the variance in the other (i.e., DOM) or to assess the statistical significance in the correlation between the two multivariate datasets. In addition, Procrustes analysis has the advantages of the application of the Procrustean association metric (i.e., residuals). Pointwise residuals indicate the difference between two different community ordinations for each sample, and are used to examine how the DOM-bacteria associations could be influenced by environmental gradients such as elevations and nutrients. The statistical significance of the Procrustes analysis (i.e.,  $M^2$ ) can then be assessed by randomly permutating the data 1000 times<sup>61</sup>. This analysis was performed using the R package `vegan` V2.4.6.

We further quantified DOM-bacteria associations at a molecular level using two different co-occurrence analyses (Box 1). First, Spearman's rank correlation coefficient  $\rho$  was calculated between the relative abundance of each molecule  $m/z$  ion and bacterial OTU (or genus). For each molecule, we then calculated the Spearman  $\rho$  difference by subtracting the mean absolute  $\rho$  value of the negative

correlations across all bacterial OTUs from the mean of the positive correlations. Larger positive and negative values indicate that molecules were more strongly positively and negatively correlated with bacterial communities, respectively. The relationships among the Spearman  $\rho$  difference, H/C and O/C were summarised using kriging interpolation with the R package `automap` V1.0.14<sup>62</sup>. Second, SparCC (Sparse Correlations for Compositional data) was applied to build DOM-bacteria bipartite networks. SparCC is a correlation method that can infer the interrelationships between DOM and bacteria for compositional data with higher accuracy<sup>35</sup> than general correlation approaches, such as Spearman's correlation, because it explicitly assumes that the underlying networks have many missing associations. We used bacterial genera rather than OTUs for bipartite network analysis because there were over 20,000 and 10,000 bacterial OTUs for Norway and China, respectively, and there are computational limits on handling such large bipartite networks for the analyses described in the next paragraph. However, using bacterial genera was reasonable as individual DOM-bacteria interactions were similar for both bacterial OTUs and genera ( $R^2 > 0.80$ ,  $P \leq 0.001$ ; Fig. S9). Similar conclusions were also obtained with either OTUs or genera when relating the pairwise distances of molecular traits with SparCC correlation coefficient  $\rho$  values among DOM molecules in Fig. 4c. To reduce type I errors in the correlation calculations created by low-occurrence genera or molecules, the majority rule was applied; that is, we retained genera or molecules that were observed in more than half of the total samples ( $\geq 75$  samples) in China or Norway. The filtered table, including 1340 and 1246 DOM molecules, and 75 and 49 bacterial genera in China and Norway, respectively, was then used for pairwise correlation calculation of DOM and bacteria using SparCC with default parameters<sup>35</sup>.

Finally, bipartite network analysis at a molecular level was performed to quantify the specialization of DOM-bacteria networks (Box 1). The specialization considers interaction abundance and is standardised to account for heterogeneity in the interaction strength and species richness, which describes the levels of "vulnerability" of DOM molecules and "generality" of bacterial taxa<sup>27</sup>. The threshold correlation for inclusion in bipartite networks was  $|\rho| = 0.30$  to exclude weak interactions and we retained the adjacent matrix with only the interactions between DOM and bacteria. We then constructed two types of interaction networks (i.e., negative and positive networks) based on negative and positive correlation coefficients (SparCC  $\rho \leq -0.30$  and  $\rho \geq 0.30$ , respectively). According to resource-consumer relationships, negative networks likely indicate the degradation of larger molecules into smaller structures, while positive networks may suggest the production of new molecules via degradation or biosynthetic processes. The SparCC  $\rho$  values were multiplied by 10,000 and rounded to integers, and the absolute values were taken for negative networks to enable the calculations of specialization indices. A separate negative and positive sub-network was obtained for each microcosm by selecting the DOM molecules and bacterial taxa in each sample based on its bacterial and DOM compositions. For the network level analysis, we calculated  $H_2'$ , a measure of specialization<sup>27</sup>, for each network:

$$H_2 = - \sum_{i=1}^i \sum_{j=1}^j (p_{ij} \ln p_{ij}) \quad (2)$$

$$H_2' = \frac{H_{2max} - H_2}{H_{2max} - H_{2min}} \quad (3)$$

where  $p_{ij} = a_{ij}/m$ , represents the proportion of interactions in a  $i \times j$  matrix.  $a_{ij}$  is the number of interactions between DOM molecule  $i$  and bacterial genus  $j$ , which is also referred as "link weight".  $m$  is the total number of interactions between all DOM molecules and bacterial genera.  $H_2'$  is the standardised  $H_2$  against the minimum ( $H_{2min}$ ) and maximum ( $H_{2max}$ ) possible for the same distribution of interaction totals. For the molecular level analysis, we calculated the specialization index Kullback-Leibler distance ( $d'$ ) for DOM molecules ( $d'_i$ ) and bacterial genera ( $d'_j$ ), which describes the levels of "vulnerability" of DOM molecules and "generality" of bacterial genera, respectively:

$$d_i = \sum_{j=1}^j \left( \frac{a_{ij}}{A_i} \ln \frac{a_{ij} m}{A_i A_j} \right) \quad (4)$$

$$d'_j = \frac{d_j - d_{min}}{d_{max} - d_{min}} \quad (5)$$

where  $A_i = \sum_{j=1}^j a_{ij}$  and  $A_j = \sum_{i=1}^i a_{ij}$ , are the total number of interactions of DOM molecule  $i$  and bacterial genus  $j$ , respectively.  $d'_i$  is the standardised  $d_i$  against the minimum ( $d_{min}$ ) and maximum ( $d_{max}$ ) possible for the same distribution of interaction totals. The equations of  $d'_j$  are analogous to  $d'_i$ , replacing  $j$  by  $i$ . Weighted means of  $d'$  for DOM were calculated for each network as the sum of the product of  $d'$  for each individual molecule  $i$  ( $d'_i$ ) and relative intensity  $I_i$  divided by the sum of all intensities  $d' = \sum (d'_i \times I_i) / \sum I_i$ . Weighted means of  $d'$  for bacteria were calculated as the sum of the  $d'$  of each individual bacterial genus  $j$  ( $d'_j$ ) and relative abundance of bacterial genus  $I_j$  divided by the sum of all abundance. All calculations were performed using the R package `FD` V1.0.12. The observed  $H_2'$  and  $d'$  values ranged from 0 (complete generalization) to 1 (complete specialization)<sup>28</sup> (Fig. S21). Specifically, elevated  $H_2'$  or  $d'$  values indicate a high degree of specialization, while lower values suggest increased generalization, that is,

higher vulnerability of DOM and/or higher generality of microbes. To directly compare the network indices across the elevations or nutrient enrichment levels, we used a null modelling approach. We standardised the three observed specialization indices ( $S_{\text{observed}}$ ; that is,  $H_2'$ ,  $d'$  of DOM, and  $d'$  of bacteria) by calculating their z-scores<sup>63</sup> using the equation:

$$z_S = (S_{\text{observed}} - \overline{S_{\text{null}}}) / (\sigma_{S_{\text{null}}}) \quad (6)$$

where  $\overline{S_{\text{null}}}$  and  $\sigma_{S_{\text{null}}}$  were, respectively, the mean and standard deviation of the null distribution of  $S$  ( $S_{\text{null}}$ ). One hundred randomised null networks were generated for each bipartite network to derive  $S_{\text{null}}$  using the *swap.web* algorithm, which keeps species richness and the number of interactions per species constant along with network connectance. This null model analysis indicates that interactions between DOM and bacteria were non-random as the observed network specialization indices were generally significantly lower than expected by chance ( $P < 0.05$ ). The obtained network was visualised using *circlize* V0.4.10<sup>64</sup> and analysed using the R package *bipartite* V2.15<sup>27</sup>.

**Statistical analyses.** We analysed how DOM features and DOM-bacteria associations varied with temperature and nutrient enrichment at both compositional- and molecular-levels, which were outlined in Fig. 2. We used the following abiotic and biotic drivers explaining variations in DOM features (i.e., alpha diversity, beta diversity, and molecular traits): water temperature, nutrient enrichment (i.e., the experimental addition of nitrate and phosphate), contemporary nutrients (i.e., TN, TP,  $\text{NO}_x^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_4^+$  and  $\text{PO}_4^{3-}$  in the sediments, and  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_4^+$  and  $\text{PO}_4^{3-}$  in the overlying water), energy supply (i.e., sediment TOC, DOC, water pH and sediment Chl *a*), and biodiversity (i.e., the species richness and the first two axes of the NMDS of bacterial community composition) (Table S2). In addition, for the response variables of DOM-bacteria associations, we used the following explanatory variables related to distal and proximal drivers (Table S1). Distal environmental drivers included climate change (i.e., water temperature), human impacts (i.e., nutrient enrichment), and contemporary nutrients (i.e., TN, TP,  $\text{NO}_x^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_4^+$  and  $\text{PO}_4^{3-}$ , and water  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{NH}_4^+$  and  $\text{PO}_4^{3-}$ ). Proximal drivers included energy supply (i.e., sediment TOC, DOC, water pH and sediment Chl *a*), biodiversity (i.e., the species richness and the first two axes of the NMDS of bacterial community composition), DOM chemodiversity (i.e., the species richness and the first two axes of the NMDS of molecular composition), and DOM molecular traits (i.e., mass, C,  $\text{Al}_{\text{Mod}}$ , DBE,  $\text{DBE}_{\text{O}}$ ,  $\text{DBE}_{\text{Al}}$ , GFE,  $\text{kdefect}_{\text{CH}_2}$ , NOSC, O/C, H/C, N/C, P/C, S/C and  $Y_{\text{met}}$ ). It should be noted that water pH could be considered to be relevant to primary productivity due to its strong positive correlation with sediment Chl *a*, but their relationships varied across elevations and nutrient levels<sup>29</sup>.

For DOM features, the relationships between nutrient enrichment and DOM richness or molecular traits were visualised with linear models for all formulae and subsets of formulae within each category of compound classes or elemental combinations across different elevations. We further tested the breakpoints or abrupt changes in DOM composition (i.e., the first axis of DCA) along the gradient of nutrient enrichment using a piecewise linear regression with the R package *segmented* V1.3.0<sup>31</sup>. The number of breakpoints was selected with the following criteria: the Bayesian Information Criteria (BIC) statistics for model selection and a maximum number of three breakpoints. We did not specify the initial nutrient values for breakpoint so that any values ranging from 0 to 36 mg N L<sup>-1</sup> would be considered. These breakpoint estimations were further supported by gradient forest analysis<sup>30</sup>, which was used to assess the DOM compositional changes and important breakpoints across multiple molecules along the gradient of nutrient enrichment. Gradient forest analysis aggregates random forest models estimated for each molecule along the nutrient gradient<sup>30</sup>, allowing us to identify non-linear changes in overall composition. This analysis produces the standardised density of splits (that is, breakpoints), that is the kernel density of splits divided by the observation density, which shows the distribution of breakpoints of each molecule regarding its abundance occurring along the nutrient gradient<sup>30</sup>. In addition, we estimated the standardised density of splits for subsets of molecules within each category of compound classes or elemental combinations across different elevations. This analysis was performed using the R packages *gradientForest* V0.1.17<sup>30</sup> and *extendedForest* V1.6.1<sup>65</sup>.

For DOM-bacteria associations, the relationships between nutrient enrichment and associations at both community and network levels were tested with linear models for all formulae and subsets of formulae within each category of compound classes or elemental combinations across different elevations.

To further evaluate the key drivers of DOM features and DOM-bacteria associations, we used variation partitioning analysis (VPA)<sup>66</sup>, random forest analysis<sup>67</sup> and structural equation modelling (SEM)<sup>36</sup>. In particular, the first analysis disentangled the important roles of bacteria from other explanatory variables, while the other non-linear and linear analyses tested the roles of molecular traits and diversity, and their interplay with environments and energy supply.

First, VPA was used to quantify the relative contributions of driver categories towards DOM features. We partitioned explanatory variables into the following driver categories: environments (that is, climate change, human impacts and contemporary nutrients), energy supply and biodiversity (Table S2). We selected explanatory variables for regression analyses by forward selection with Akaike

information criterion (AIC)<sup>68</sup>. We also quantified the relative contributions of driver categories for subsets of molecules within each category of compound classes or elemental combinations. VPA was performed with R package *vegan* V2.4.6<sup>69</sup>.

Second, random forest analysis was conducted to identify the relative importance of environment variables, energy supply, bacterial diversity and DOM molecular drivers on specialization  $H_2'$ . The importance of each predictor variable was determined by evaluating the decrease in prediction accuracy (that is, increase in the mean square error between observations and out-of-bag predictions) when the data for that predictor were randomly permuted. The accuracy importance measure was computed for each tree and averaged over the forest (2000 trees). More details on this method were described in previous literature<sup>70</sup>. This analysis was conducted using the R package *randomForestSRC* V2.8.0<sup>71,72</sup>.

Third, SEM was used to explore how specialization  $H_2'$  is interactively influenced by global changes (that is, temperature and nutrient enrichment), diversity and molecular traits. The approach begins by hypothesising the underlying structure of causal links as shown in Fig. 6a. Then, the model is translated into regression equations, and these equations are evaluated against the data to test the hypothesised links. Through this process, SEM provides an understanding of direct and indirect links of climate change and human impacts on  $H_2'$ . Before modelling, all variables in the SEMs were z-score transformed to allow comparisons among multiple predictors and models. Similar to previous studies<sup>73</sup>, we used composite variables to account for the collective effects of climate change, human impacts, contemporary nutrients, energy supply, biodiversity, chemodiversity and molecular traits, and the candidate observed indicators are given in Table S1. The indicators for each composite were selected based on the multiple regressions for  $H_2'$  (Table S3). Based on all the hypothesised links among composite variables (that is, full model; Fig. 6a), we examined all alternative models using AIC and overall model fit statistics<sup>74</sup>. We chose the final model to report as that with the lowest AIC value from models with a non-significant  $\chi^2$  test ( $P > 0.05$ ), which tests whether the model structure differs from the observed data, high comparative fit index (CFI > 0.95) and low standardised root mean squared residual (SRMR < 0.05) (Table S4). We implemented the SEMs using R package *lavaan* V.0.5.23<sup>75</sup>.

**Predictions of DOM-bacteria network specialization in Taihu Lake.** Using the parameter estimates obtained from SEM fitted to the bipartite networks in sub-tropical China, we estimated spatiotemporal variation of DOM-bacteria network specialization in Taihu Lake based on the direct and indirect effects of climate change and eutrophication via the proximal drivers. We first formulated five linear equations to predict the values of contemporary nutrients ( $P_{\text{nut}}$ ), energy supply ( $P_{\text{energy}}$ ), biodiversity ( $P_{\text{biodiv}}$ ), chemodiversity ( $P_{\text{chemodiv}}$ ) and molecular traits ( $P_{\text{trait}}$ ) based on climate and eutrophication drivers:

$$P_{\text{nut}} = \lambda_{\text{nut,temp}} \times X_T + \lambda_{\text{nut,N}} \times X_N \quad (7)$$

$$P_{\text{energy}} = \lambda_{\text{energy,temp}} \times X_T + \lambda_{\text{energy,N}} \times X_N + \lambda_{\text{energy,nut}} \times P_{\text{nut}} \quad (8)$$

$$P_{\text{biodiv}} = \lambda_{\text{biodiv,temp}} \times X_T + \lambda_{\text{biodiv,N}} \times X_N + \lambda_{\text{biodiv,nut}} \times P_{\text{nut}} + \lambda_{\text{biodiv,energy}} \times P_{\text{energy}} \quad (9)$$

$$P_{\text{chemodiv}} = \lambda_{\text{chemodiv,temp}} \times X_T + \lambda_{\text{chemodiv,N}} \times X_N + \lambda_{\text{chemodiv,nut}} \times P_{\text{nut}} + \lambda_{\text{chemodiv,energy}} \times P_{\text{energy}} \quad (10)$$

$$P_{\text{trait}} = \lambda_{\text{trait,temp}} \times X_T + \lambda_{\text{trait,N}} \times X_N + \lambda_{\text{trait,nut}} \times P_{\text{nut}} + \lambda_{\text{trait,energy}} \times P_{\text{energy}} + \lambda_{\text{trait,biodiv}} \times P_{\text{biodiv}} + \lambda_{\text{trait,chemodiv}} \times P_{\text{chemodiv}} \quad (11)$$

where  $X_T$  and  $X_N$  were water temperature and total nitrogen, respectively, for the 32 sites across the whole Taihu Lake (Fig. S19a). The abbreviations of path coefficients ( $\lambda$ ) are detailed in Table S5.

Similarly, we calculated the specialization of DOM-bacteria networks ( $Y_{H_2}$ ) using a linear equation:

$$Y_{H_2} = \lambda_{H_2,temp} \times X_T + \lambda_{H_2,N} \times X_N + \lambda_{H_2,nut} \times P_{\text{nut}} + \lambda_{H_2,energy} \times P_{\text{energy}} + \lambda_{H_2,biodiv} \times P_{\text{biodiv}} + \lambda_{H_2,chemodiv} \times P_{\text{chemodiv}} + \lambda_{H_2,trait} \times P_{\text{trait}} \quad (12)$$

We used the predicted values for contemporary nutrients, energy supply, biodiversity, chemodiversity and molecular traits in the overall prediction model to account for the indirect effects of water temperature and total nitrogen on specialization. The models were calculated with a yearly time step based on the annual means of water temperature and total nitrogen for each site during 2007–2018. The temporal changes in specialization were calculated using 2007 as a baseline to which all predictions were compared.

The above predictions aimed to apply our EDTiA framework to estimate changes in DOM-bacteria network specialization under temperature change and eutrophication in Taihu Lake, and potential uncertainties in the estimated specialization should however be noted as follows. First, local environmental variation (e.g. N/P ratio changes) and different microbial species pools between our field microcosms and natural lake sediments would likely influence the accuracy of predictions. Second, spatial and temporal heterogeneity of sediments would influence local environments and the composition of both DOM and

microbes and thus the projection of estimates across Taihu Lake. Third, the transferability and extrapolation of SEM models to Taihu Lake would be one of the difficulties in prediction practices. We thus selected the SEM models in China rather than Norway for more similar climatic conditions to the target lake. The annual mean water temperatures in Taihu Lake were covered by the temperature variations across the elevations between 2286 and 3822 m a.s.l. in Laojun Mountain, and the annual mean total nitrogen fell into the gradient of nutrient concentrations between 0 and 36 mg N L<sup>-1</sup>. Finally, lake management such as mechanical removal of algae would affect energy supply and consequently prediction accuracy.

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

### Data availability

The microbial sequences and meta data generated in this study have been deposited in the MG-RAST database under accession code 17710. The monthly monitoring environmental data of Taihu Lake is available under restricted access according to data management policy and the access could be obtained from Taihu Laboratory for Lake Ecosystem Research (<http://thl.cern.ac.cn/>) or the corresponding author upon reasonable request. The other data are available under restricted access due to the authors' continuing projects of field experiments on global mountainsides, and the access can be obtained from the corresponding author upon reasonable request.

### Code availability

The R codes for specialization metrics are available at <https://doi.org/10.5281/zenodo.6592491>, and at <http://github.com/jianjunwang/iDOM>.

Received: 30 June 2021; Accepted: 9 June 2022;

Published online: 23 June 2022

### References

- Dittmar, T. & Stubbins, A. in *Treatise on Geochemistry* 2nd edn, 125–156 (Elsevier, 2014).
- Zhang, C. et al. Evolving paradigms in biological carbon cycling in the ocean. *Nat. Sci. Rev.* **5**, 481–499 (2018).
- Kujawinski, E. B. The impact of microbial metabolism on marine dissolved organic matter. *Annu. Rev. Mar. Sci.* **3**, 567–599 (2011).
- Jiao, N. et al. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* **8**, 593–599 (2010).
- Lennon, J. & Pfaff, L. Source and supply of terrestrial organic matter affects aquatic microbial metabolism. *Aquat. Micro. Ecol.* **39**, 107–119 (2005).
- Berggren, M. et al. Lake secondary production fueled by rapid transfer of low molecular weight organic carbon from terrestrial sources to aquatic consumers. *Ecol. Lett.* **13**, 870–880 (2010).
- Cherif, M. & Loreau, M. Stoichiometric constraints on resource use, competitive interactions, and elemental cycling in microbial decomposers. *Am. Nat.* **169**, 709–724 (2007).
- Kellerman, A. M., Kothawala, D. N., Dittmar, T. & Tranvik, L. J. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat. Geosci.* **8**, 454–457 (2015).
- Roth, V.-N. et al. Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).
- Judd, K. E., Crump, B. C. & Kling, G. W. Variation in dissolved organic matter controls bacterial production and community composition. *Ecology* **87**, 2068–2079 (2006).
- Orland, C., Yakimovich, K. M., Mykityczuk, N. C. S., Basiliko, N. & Tanentzap, A. J. Think global, act local: The small-scale environment mainly influences microbial community development and function in lake sediment. *Limnol. Oceanogr.* **65**, S88–S100 (2020).
- Wang, J. et al. Embracing mountain microbiome and ecosystem functions under global change. *New Phytol.* **234**, 1987–2002 (2022).
- Osterholz, H. et al. Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *ISME J.* **10**, 1717–1730 (2016).
- Tanentzap, A. J. et al. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc. Natl Acad. Sci. USA* **116**, 24689 (2019).
- Danczak, R. E. et al. Using metacommunity ecology to understand environmental metabolomes. *Nat. Commun.* **11**, 6369 (2020).
- Li, H.-Y. et al. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. *Microbiome* **6**, 187 (2018).
- Sarmento, H. & Gasol, J. M. Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ. Microbiol.* **14**, 2348–2360 (2012).
- Baines, S. B. & Pace, M. L. The production of dissolved organic matter by phytoplankton and its importance to bacteria: Patterns across marine and freshwater systems. *Limnol. Oceanogr.* **36**, 1078–1090 (1991).
- Herzprung, P. et al. Improved understanding of dissolved organic matter processing in freshwater using complementary experimental and machine learning approaches. *Environ. Sci. Technol.* **54**, 13556–13565 (2020).
- Sarmento, H., Morana, C. & Gasol, J. M. Bacterioplankton niche partitioning in the use of phytoplankton-derived dissolved organic carbon: quantity is more important than quality. *ISME J.* **10**, 2582–2592 (2016).
- Wagner, S. et al. Soothsaying DOM: a current perspective on the future of oceanic dissolved organic carbon. *Front. Marine Sci.* **7**, 341 (2020).
- Treseder, K. K., Kivlin, S. N. & Hawkes, C. V. Evolutionary trade-offs among decomposers determine responses to nitrogen enrichment. *Ecol. Lett.* **14**, 933–938 (2011).
- Finke, D. L. & Snyder, W. E. Niche partitioning increases resource exploitation by diverse communities. *Science* **321**, 1488 (2008).
- Muscarella, M. E., Boot, C. M., Broeckling, C. D. & Lennon, J. T. Resource heterogeneity structures aquatic bacterial communities. *ISME J.* **13**, 2183–2195 (2019).
- D'Andrilli, J., Cooper, W. T., Foreman, C. M. & Marshall, A. G. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **29**, 2385–2401 (2015).
- Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).
- Dormann, C. F., Frund, J., Bluthgen, N. & Gruber, B. Indices, graphs and null models: analyzing bipartite ecological. *Netw. Open Ecol. J.* **2**, 7–24 (2009).
- Bluthgen, N., Menzel, F. & Bluthgen, N. Measuring specialization in species interaction networks. *BMC Ecol.* **6**, 9 (2006).
- Wang, J., Pan, F., Soininen, J., Heino, J. & Shen, J. Nutrient enrichment modifies temperature-biodiversity relationships in large-scale field experiments. *Nat. Commun.* **7**, 13960 (2016).
- Ellis, N., Smith, S. J. & Pitcher, C. R. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).
- Muggeo, V. Segmented: an R package to fit regression models with broken-line relationships. *R. N.* **8**, 20–25 (2008).
- Gower, J. C. Generalized procrustes analysis. *Psychometrika* **40**, 33–51 (1975).
- Hurley, J. & Cattell, R. The PROCUSTES program: producing direct rotation to test a hypothesized factor structure. *Behav. Sci.* **7**, 258–262 (2007).
- Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, 2012).
- Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
- Grace, J. B. et al. Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* **3**, art73 (2012).
- Sinha, E., Michalak, A. M. & Balaji, V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* **357**, 405 (2017).
- Anderson, N. J., Bennion, H. & Lotter, A. F. Lake eutrophication and its implications for organic carbon sequestration in Europe. *Glob. Change Biol.* **20**, 2741–2751 (2014).
- Kothawala, D. N., Kellerman, A. M., Catalan, N. & Tranvik, L. J. Organic matter degradation across ecosystem boundaries: the need for a unified conceptualization. *Trends Ecol. Evol.* **36**, 113–122 (2021).
- Kellerman, A. M. et al. Unifying concepts linking dissolved organic matter composition to persistence in aquatic ecosystems. *Environ. Sci. Technol.* **52**, 2538–2548 (2018).
- Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
- Jiao, N. et al. Correcting a major error in assessing organic carbon pollution in natural waters. *Sci. Adv.* **7**, eabc7318 (2021).
- Wieder, W. R. et al. Explicitly representing soil microbial processes in Earth system models. *Glob. Biogeochem. Cy* **29**, 1782–1800 (2015).
- Chadburn, S. E. et al. Modeled microbial dynamics explain the apparent temperature sensitivity of wetland methane emissions. *Global Biogeochem. Cy* **34**, e2020GB006678 (2020).
- Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Choi, J. H. et al. Influence of biogenic organics on the chemical composition of arctic aerosols. *Glob. Biogeochem. Cy* **33**, 1238–1250 (2019).
- Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Meth.* **6**, 230–235 (2008).
- Tolić, N. et al. Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Anal. Chem.* **89**, 12659–12665 (2017).

49. Kujawinski, E. B. & Behn, M. D. Automated analysis of electrospray ionization Fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* **78**, 4363–4373 (2006).
50. Koch, B. P., Dittmar, T., Witt, M. & Kattner, G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **79**, 1758–1763 (2007).
51. Kim, S., Kramer, R. W. & Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344 (2003).
52. Patriarca, C. et al. Investigating the ionization of dissolved organic matter by electrospray. *Anal. Chem.* **92**, 14210–14218 (2020).
53. Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier Scientific, 1998).
54. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
55. Koch, B. P. & Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **30**, 250–250 (2016).
56. LaRowe, D. E. & Van Cappellen, P. Degradation of natural organic matter: a thermodynamic analysis. *Geochim. Cosmochim. Acta* **75**, 2030–2042 (2011).
57. Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G. & Qian, K. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* **73**, 4676–4681 (2001).
58. Song, H.-S. et al. Representing organic matter thermodynamics in biogeochemical reactions via substrate-explicit modeling. *Front. Microbiol.* **11**, 531756 (2020).
59. Schneidman-Duhovny, D. et al. ftmsRanalysis: an R package for exploratory data analysis and interactive visualization of FT-MS data. *PLoS Comput. Biol.* **16**, e1007654 (2020).
60. Lavorel, S. et al. Assessing functional diversity in the field—methodology matters! *Funct. Ecol.* **22**, 134–147 (2007).
61. Peres-Neto, P. R. & Jackson, D. A. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**, 169–178 (2001).
62. Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W. & Heuvelink, G. B. M. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci.* **35**, 1711–1721 (2009).
63. Dormann, C. F., Strauss, R. & Peres-Neto, P. A method for detecting modules in quantitative bipartite networks. *Methods Ecol. Evol.* **5**, 90–98 (2014).
64. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. *circlize* Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
65. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R. N.* **2**, 18–22 (2002).
66. Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).
67. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
68. Miller, J. K. & Farr, S. D. Bimultivariate redundancy: a comprehensive measure of interbattery relationship. *Multivar. Behav. Res.* **6**, 313–324 (1971).
69. Oksanen, J. et al. *vegan: Community Ecology Package*. CRAN R package (2017).
70. Hu, A. et al. Mountain biodiversity and ecosystem functions: interplay between geology and contemporary environments. *ISME J.* **14**, 931–944 (2020).
71. Ishwaran, H. & Kogalur, U. B. Random survival forests for R. *R. N.* **7**, 25–31 (2007).
72. Ishwaran, H. & Kogalur, U. B. *Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)* (2019).
73. Grace, J. B. et al. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529**, 390 (2016).
74. Grace, J. B., Anderson, T. M., Olff, H. & Scheiner, S. M. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* **80**, 67–87 (2010).
75. Rosseel, Y. lavaan: an R package for structural equation Modeling. *J. Stat. Softw.* **48**, 1–36 (2012).

## Acknowledgements

We appreciate Chengyan Zhang, Lizhou Dai, Fanfan Meng, Yongqiang Zhou, Hao Wu, Shaojuan Du and Feiyan Pan for field sampling, lab/data analyses and figures, Guoping Zhao, Nianzhi Jiao, Yahai Lu, Hailiang Dong, Chuanlun Zhang, Qinglong Wu, Ganlin Zhang, Suming Wang and Nik Cunniffe for valuable suggestions on our preliminary draft, and Thorsten Dittmar for a key introduction. We thank Taihu Laboratory for Lake Ecosystem Research for kindly providing long-term data of Taihu Lake. This study was supported by National Natural Science Foundation of China (91851117 to J.W., 42077052 to A.H.), the National Key Research and Development Program of China (2017YFA0605203 to J.W.), CAS Key Research Program of Frontier Sciences (QYZDB-SSW-DQC043 to J.W.), National Natural Science Foundation of China (41871048 to J.W. and 41273088 to J.W.), Independent Scientific Project of Jiangsu Provincial Academy of Environmental Sciences (C21047 to J.W.) and Research Program of Sino-Africa Joint Research Center, Chinese Academy of Sciences (151542KYSB20210007 to J.W.). K.J. was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) (NRF-2021M1A5A1075510 to K.J.) and KBSI (C140440 to K.J.) grants. A.J.T. was supported by H2020 ERC Grant (sEEingDOM 804673 to A.J.T.). J.T.L. was supported by National Science Foundation (DEB-1442246 to J.T.L., 1934554 to J.T.L.), US Army Research Office Grant (W911NF-14-1-0411 to J.T.L.), and the National Aeronautics and Space Administration (80NSSC20K0618 to J.T.L.).

## Author contributions

J.W. conceived the idea. J.W. carried out the field trips and provided the physicochemical and biological data. J.L., M.C. and K.J. analysed the DOM. A.H. and J.W. performed the statistical analyses. A.H. wrote the first draft of the manuscript. A.H. and J.W. finished the manuscript with the comments from A.J.T., J.T.L., J.S., K.J., Y.L., X.L. and Y.Z. All authors contributed to the intellectual development of this study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-31251-1>.

**Correspondence** and requests for materials should be addressed to Jianjun Wang.

**Peer review information** *Nature Communications* thanks Helen Glanville and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022