Quantifying microbial associations of dissolved organic matter under global change

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

[1] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academic of Sciences, Nanjing 210008, China

[2] College of Resources and Environment, Hunan Agricultural University, Changsha 410128, China

[3] Bio-Chemical Analysis Team, Korea Basic Science Institute, Cheongju 28119, South Korea

[4] Ecosystems and Global Change Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[5] Nanchang Institute of Technology, Nanchang 330099, China

[6] Department of Biology, Indiana University, Bloomington IN 47405, USA

[7] Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China.

[8] University of Chinese Academy of Sciences, Beijing 100049, China

[9] Department of Geosciences and Geography, University of Helsinki, Helsinki FIN-00014, Finland

[10] State Key Laboratory for Mineral Deposits Research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210093, China

[11] School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China

[*] Correspondence: jjwang@niglas.ac.cn

Running title: DOM-microbe associations under global change

Abstract

Microbes play a critical role in regulating the size, composition, and turnover of dissolved organic matter (DOM), which is one of the largest pools of carbon in aquatic ecosystems. Global change may alter DOM-microbe associations with implications for biogeochemical cycles, although disentangling these complex interactions remains a major challenge. Here we develop a framework called Energy-Diversity-Trait integrative Analysis (EDTiA) to examine the associations between DOM and bacteria along temperature and nutrient gradients in a manipulative field experiment on mountainsides in contrasting subarctic and subtropical climates. In both study regions, the chemical composition of DOM correlated with bacterial communities, and was primarily controlled by nutrients and to a lesser degree by temperature. At a molecular-level, DOM-bacteria associations depended strongly on the molecular traits of DOM, with negative associations indicative of decomposition as molecules are more biolabile. Using bipartite networks, we further demonstrated that negative associations were more specialized than positive associations indicative of DOM production. Nutrient enrichment promoted specialization of positive associations, but decreased specialization of negative associations particularly at warmer temperatures in subtropical climate. These global change drivers influenced specialization of negative associations most strongly via molecular traits, while both molecular traits and bacterial diversity similarly affected positive associations. Together, our framework provides a quantitative approach to understand DOM-microbe associations and wider carbon cycling across scales under global change.

Introduction

Dissolved organic matter (DOM), one of the largest pools of carbon in aquatic ecosystems [1], is intimately interlinked with the metabolic processes of complex microbial communities [2]. Microbial consortia generate "chemodiversity" in the DOM pool by degrading larger molecules into smaller molecules and by synthesizing more refractory compounds from labile substrates [3]. These basic processes together lead to the emergence of molecular traits of DOM including chemical structure, stoichiometry, oxidation state, and bioavailability [4-6] that directly determine its environmental persistence [7, 8]. DOM, as a carbon source for microbial metabolism, also influences the diversity, structure, and functioning of microbial communities via decomposition and biosynthetic processes [9-13]. The resulting resource-consumer relationships can now be characterised in both aquatic [14-16] and terrestrial [17] ecosystems owing to recent advances in ultrahigh-resolution mass spectrometry and high-throughput sequencing. Despite the availability of these technologies, little is known about how DOM-microbe associations can be quantified in nature, and are interactively and independently affected by global change drivers, such as elevated temperatures and eutrophication.

The effects of global change on DOM-microbe associations can be viewed through three proximal controls (Fig. 1). First, energy supply, such as primary productivity, represents the major source of DOM that supports microbial metabolism [18-20]. In particular, elevated temperature and nutrient inputs can stimulate primary productivity in ways that influences the composition and availability of organic matter [21, 22], but this process also indirectly influence DOM-microbe associations by controlling their diversity and traits [22-25]. Second, diversity can generally beget diversity. For example, an increase in the diversity of DOM promotes microbial diversity, and vice versa [15], which should be reflected as signatures in resource-consumer relationships. Such patterns may arise because resource diversity promotes microbial specialization during biochemical transformations by creating more unique resource niches for consumers to partition [26, 27]. Likewise, higher microbial diversity provides more metabolic pathways to decompose and produce molecules, which influences the vulnerability of DOM to degradation [3]. Third, DOM-microbe associations depend on the molecular traits of DOM,

such as its bioavailability, measured with H/C ratios of individual molecules [28], and microbial traits like life history (i.e., *r*- versus *K*-selection) [29] and resource acquisition (i.e., generalists versus specialists) [27].

To integrate the three proximal controls to examine how DOM-microbe associations vary under global change, we developed a framework called Energy-Diversity-Trait integrative Analysis (EDTiA) (Fig. 1). EDTiA relies on the construction of bipartite networks [30] to quantify the specialization between organic molecules and microbial taxa. These networks are investigated using measures of entropy such as the H_2' index [31], which quantify resource-consumer relationships at an ecosystem-level. For example, elevated H_2' values convey that there is a high degree of specialization between DOM and microbes [31], where in the extreme example, one bacterial taxon consumes or produces a single DOM molecule. By contrast, lower H_2' reflects a more generalized bipartite network where different DOM molecules can be used by a large range of bacterial taxa. Furthermore, EDTiA allows for the integration of global change drivers to explore their relative importance of proximal controls on the specialization of DOM-microbe associations (Fig. 1).

We therefore used the EDTiA framework to test how associations between DOM and bacteria were jointly influenced by temperature and nutrient loading in a manipulative field experiment on subtropical and subarctic mountainsides in China and Norway [32]. This macroecological approach involved creating microcosms with consistent initial DOM composition but different locally colonised microbial communities and newly produced endogenous DOM. Briefly, we selected five locations with different elevations on each mountainside that spanned a mean annual temperature gradient of 4.2-12.9°C in China and -2.9-0.7 °C in Norway. We established 300 sterile aquatic microcosms composed of natural lake sediments and artificial lake water, which included ten nutrient levels at each elevation. The sediments originated from Taihu Lake, a large eutrophic shallow lake in China, and were added to each microcosm after sterilisation to ensure identical initial DOM supply and composition. Microcosms were left in the field for one month allowing airborne bacteria to colonise, and sediment bacteria were examined using high-throughput sequencing of 16S rRNA genes [32]. Additionally, we

applied ultrahigh-resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to examine sediment DOM features, such as chemodiversity and molecular traits.

Our study addresses three questions: (1) How do associations between chemodiversity and microbial biodiversity respond to temperature and nutrient enrichment? (2) How does the specialization of molecular-level associations between DOM and microbes vary along temperature and nutrient gradients? (3) How is the specialization interactively and independently influenced by temperature and nutrient enrichment via the three proximal controls? Results from our study will help advance biogeochemical modeling and improve predictions about carbon turnover along with feedbacks based on resource-based constraints on microbial diversity.

Results and Discussion

(1) DOM features and their microbial associations at a compositional-level

The diversity and molecular traits of DOM were strongly controlled by nutrient enrichment but less by temperature in both mountainsides (Figs. S2-4). Nutrient enrichment generally promoted molecular richness in both regions when all molecular components were considered (Figs. 2a, S5). Using piecewise regression [33] and gradient forest analysis [34], we identified breakpoints in molecular composition that mostly occurred between 1.80 and 4.05 mg N L^{-1} along the nutrient gradient for all molecules at each elevation (Figs. 2a, S6). The effects of nutrient enrichment on molecular traits, however, varied between the two ecoregions (Figs. 2a, S5, S7). 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. 2a, S5c). The ratio remained consistently higher (≥ 1.5) across all nutrient levels in Norway (Figs. 2a, S5c). Given the identical DOM composition in our study initially, 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 highlights the need for a more mechanistic approach offered by the EDTiA framework.

DOM composition was strongly associated with bacteria 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 (1.0 to 26.7% of explained variation) indicated that these variables also indirectly influenced the associations between DOM and bacteria (Figs. S8-10). These DOM-microbe associations were also supported by a Procrustes analysis [35, 36], which revealed that more similar mixtures of molecules were related to more similar bacterial communities ($M^2 = 0.701$, $P \leq 0.001$; Fig. 2b), and their associations varied with temperature (that is, elevation) 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 temperatures in China (Fig. S11). In Norway, the differences were always lower, on average, and did not vary with temperature (Fig. S11). Nutrient enrichment similarly influenced the correlations between DOM molecular and bacterial compositions estimated from both alpha and beta diversity (Fig. 2c), and these correlations also varied with nutrients for individual compound classes or elemental combinations (Figs. S12, S13). Interestingly, the coordinated compositional changes in DOM and bacteria, measured by the correlation between beta diversities, increased more strongly with nutrient enrichment in Norway than in China, especially at low nutrient levels beneath 1.80 mg N L^{-1} (Figs. 2c, S13).

(2) Networks between DOM and bacteria at a molecular-level

To quantify the associations between DOM and bacteria further at a molecular-level, we first correlated the relative abundance of DOM molecules and bacterial taxa. According to resource-consumer relationships, negative associations likely indicate the degradation of larger molecules into smaller structures, while positive associations may relate to the production of new molecules via degradation or biosynthetic processes. We found that the distribution of negative and positive correlations between DOM molecules

and bacterial operational taxonomic units (OTUs) depended strongly on molecular traits. For example, more labile molecules, such as those with H/C ≥ 1.5, were more likely to show negative Spearman's correlation coefficients ρ with individual OTUs ($P \leq 0.05$), whereas more recalcitrant molecules (H/C < 1.5) generally showed more positive correlations ($P \leq 0.05$), especially in Norway (Fig. S14). These findings were even more clearly supported by the differences between the mean of the positive and negative ρ values for each molecule (Figs. 3a, S15). 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 were the outcome of degradation processes, while ρ 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. 3a).

Subsequently, we quantified DOM-microbe associations 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) [37]. SparCC relies on algorithms for sparse neighborhood and inverse covariance selection, and can infer correlations with a high degree of accuracy under these conditions [37]. In total, there were 6,916 and 8,409 interactions for negative and positive networks (|SparCC ρ| ≥ 0.3), respectively, in China, and 1,313 and 2,888 negative and positive interactions, respectively, in Norway (Fig. 3b). The weighted mean of the percentage of SparCC ρ values that were strongly negative ($P \leq 0.05$) increased towards high nutrient levels, with the reverse pattern for positive networks, almost exclusively in China (Fig. S16). Such patterns were consistent with the weighted mean SparCC ρ in China (Fig. S16).

The negative and positive interaction networks strongly depended on molecular traits, which was further supported by three observations. First, negative and positive networks were associated with different molecule groups categorised by hierarchical cluster analysis based on the 16 molecular traits (Figs. 3b, S17). In China, negative interactions were dominant between molecule clusters 4 and 5, which were largely comprised of recalcitrant molecules with a H/C of < 1.5 (Fig. S17), and bacteria in the phyla Proteobacteria, Bacteroidetes, or Firmicutes (Fig. 3b). The positive interactions

were mostly linked to clusters 1 and 3 (Fig. 3b), which mostly represented labile molecules with a H/C of ≥ 1.5 (Fig. S17). In Norway, molecule cluster 4 was mainly negatively linked to Firmicutes and positively linked to α- and β-Proteobacteria (Fig. 3b). Second, molecules generally covaried more similarly with microbes as they shared more similar traits. For example, we detected statistically significant correlations between the pairwise Gower distances [38] of the traits and SparCC ρ values of DOM molecules in each region (Mantel test, $P \leq 0.001$; Fig. 3c). Third, molecular traits were more strongly correlated with SparCC ρ in the negative than positive interaction networks for all molecules (Fig. 3c), which was also true for most of the networks when considering compound classes or elemental combinations (Fig. 3d). These correlations, consistent at the OTU level (Fig. 3c), 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 [31]. We also calculated specialization d' indices for individual DOM molecules and bacterial genera [31]. Elevated H_2' or d' values indicate a high degree of specialization, while lower values suggest increased generalization. We found that networks that were more specialized in the negative associations between DOM and bacteria (i.e., higher H_2' values) corresponded with more specialized communities of DOM molecules (i.e., higher weighted mean d'; Figs. S18, S19) [31]. For positive networks, H_2' values mirrored those of d' for both DOM and bacteria (Figs. S18, S19). These results 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. 4a, S20), indicating a higher degree of specialization in the decomposition than production processes of microbes. Copiotrophs may have a high substrate specificity for labile resources as compared with oligotrophs [39], which have multiple metabolic pathways for resource acquisition of complex organic matter and hence lower specialization [40]. 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. 4a), suggesting more specialized decomposition (i.e., negative networks) and thus potentially more degradable DOM 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. 4a), which was particularly true when considering only recalcitrant components, such as lignin and CHNO (Fig. S21). Compared to Norway, nutrient enrichment increased the H_2' of positive interactions relatively more at lower elevations in China (Fig. 4a). Nutrient enrichment at the warmer temperatures in the subtropical region could thus contribute to the greater recalcitrance of DOM by reducing the specialization of decomposition (i.e., negative networks) and resulting in more specialized production of molecules (i.e., positive networks).

(3) Drivers of DOM-microbe associations

We explored the following distal and proximal controls on negative and positive DOM-microbe networks under 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 the DOM molecular traits (Table S1). In addition to bacterial diversity and chemodiversity, molecular traits strongly correlated with H_2' and influenced it through hypothesised casual relationships in structural equation models (SEM) [41] (Fig. S1).

The importance of molecular traits was supported by Pearson correlations (Fig. S22), multiple regression models (Fig. S23) and random forest analyses [42] (Fig. 4b). For the negative networks, H_2' showed the highest Pearson correlation coefficient of $r = 0.77$ with molecular composition ($P \leq 0.001$), followed by molecular richness ($r = -0.76$, $P \leq 0.001$) and N/P ratio ($r = 0.76$, $P \leq 0.001$, Fig. S22). In contrast, H_2' was less correlated with molecular traits for the positive networks (Fig. S22). Multiple regression models

revealed that, for negative and positive networks in China, there were statistically significant ($P \leq 0.01$) improvements in the explained variances of models by between 6.2% and 9.1% from including either diversity or molecular traits (Fig. S23). These improvements were larger for the negative interaction networks in Norway (Fig. S23). These effects of diversity and molecular traits were further supported by random forest analyses. Diversity and molecular traits improved the predictive power of models of H_2' by 7.9-26.1% and 2.1-14.8%, respectively, and again, most strongly for the negative interactions in both regions (Fig. S23). Furthermore, 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. 4b).

We also used SEM to test the hypothesised effects of two global change drivers, climate change and human impacts, on the specialization of DOM-bacteria networks. We compared these effects to other drivers like contemporary nutrients, energy supply, biodiversity, chemodiversity and molecular traits (Fig. S1). The SEM results strongly indicated that there were different constraints on DOM-microbe specialization between negative and positive interaction networks. For the negative networks, both global change drivers strongly influenced H_2' through indirect effects on energy supply and molecular traits, especially in China (Figs. 5, S24). 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 networks in China (Fig. 5a). 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. 5b, S24). For the positive networks, 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. 5, S24).

(4) Implications

The factors that control microbial processing of DOM composition, and consequently its degradation, remain challenging to discern [43], yet are critical for predicting carbon cycling under global change scenarios. We found that associations between DOM and microbial decomposers depended on universal drivers of ecosystem functioning, such as energy supply [21, 22], both DOM and microbial compositions [8, 15], and molecular traits [7, 28]. The EDTiA framework we developed provides a unified approach to identify when each of these different 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, we found that molecular traits generally helped shape DOM-microbe networks across contrasting climatic zones, especially the negative networks indicative of degradation processes. Although molecular traits are well known to be associated with DOM persistence or vulnerability to degradation [7, 44], their influence on the underlying biological mechanisms has remained poorly understood. Our results advance this work by demonstrating when the specialization of DOM-microbe associations changes with molecular traits, and by providing predictions of how specialization might vary under global change.

We found that temperature and eutrophication can change DOM-microbe associations by shifting the three proximal drivers, namely energy, diversity, and traits. For positive bipartite networks, nutrient enrichment generally increased the specialization of DOM-microbe associations, 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 d' (Fig. S19). By contrast, both temperature and nutrient enrichment reduced specialization in the negative networks, primarily via changing molecular traits and energy supply. Decomposition processes associated with negative networks may depend more on whether molecules contain structures that resist degradation [7], especially in the absence of temperature limitation [45]. At higher temperatures, such as in subtropical China, energy to degrade these molecules may become more limiting [45]. We also found that the importance of these distal drivers of climate change and human impacts varied between biomes. For instance, both elevated

319 temperature and eutrophication reduced the specialization of negative DOM-microbe
320 networks indicative of decomposition processes in subtropical China, but these two
321 drivers were less important in subarctic Norway. As their indirect effects via microbial
322 composition varied between biomes, these responses may reflect differences in the
323 biological traits of communities. Future studies with metagenomics could offer a
324 powerful complement to test how microbial traits vary with DOM traits.

325 As inland water worldwide continues to undergo changes in climate [46] and trophic
326 state [47], our approaches could be applied to predict changes in how microbes degrade and
327 produce DOM. For instance, since hyper-eutrophication occurred in Taihu Lake in May
328 2007, total nitrogen has been reduced by a mean (± SD) of 1.24 (± 1.41) mg L^{-1} with
329 strong lake management (Figs. S25, S26). Based on the estimated direct and indirect
330 effects of distal controls in the SEM fitted to the Chinese data (Fig. S1), this
331 oligotrophication, combined with a mean decrease in water temperature of 0.20 (±
332 0.87) °C between 2007 and 2018, was predicted to change the specialization of DOM-
333 microbe associations. Specifically, H_2' changed by +0.65 (± 0.58) and -0.65 (± 0.46) for
334 negative and positive networks, respectively, over this period (Fig. 6a). The greatest
335 changes happened in the most eutrophic part of the lake, including the northwestern
336 lakeshore and the northern Zhushan and Meiliang Bays (Figs. 6b, S27). Although our
337 predictions ignored detailed spatiotemporal environmental variations as used to
338 parameterize the SEM models, they do illustrate the potential to upscale our predictions
339 in real-world settings. This understanding, facilitated through our EDTiA framework,
340 could provide the first steps for improving Earth system models of global biogeochemical
341 cycles [48]. More generally, our work shows how the molecular traits of DOM will control
342 the responses of DOM-microbe networks and their associated biogeochemical cycles in a
343 changing world.

Acknowledgements

We appreciate C.Y. Zhang, L.Z. Dai and F.Y. Pan for field sampling and lab analyses, and T. Dittmar and C.L. Zhang for valuable suggestions. We thank Taihu Laboratory for Lake Ecosystem Research for providing long-term data of Taihu Lake. This study was supported by National Natural Science Foundation of China (91851117), CAS Key Research Program of Frontier Sciences (QYZDB-SSW-DQC043), and National Natural Science Foundation of China (42077052, 41871048). KJ was supported by Korea Basic Science Institute grants (C140222, C140444). AJT was supported by H2020 ERC Grant (sEEIngDOM 804673). JTL was supported by National Science Foundation (DEB-1442246, 1934554), US Army Research Office Grant (W911NF-14-1-0411), and the National Aeronautics and Space Administration (80NSSC20K0618).

Conflict of Interest

The authors declare no conflict of interest.

Author contributions

JW conceived the idea. JW carried out the field trips and provided the physiochemical and biological data. JL, MC and KJ analysed the DOM. AH and JW performed the statistical analyses. AH wrote the first draft of the manuscript. AH and JW finished the manuscript with the comments from AJT, JTL, JS, KJ, YL and XL. All authors contributed to the intellectual development of this study.

Data availability

Bacterial sequences and environmental data are available in Wang *et al* (2016) [32]. Other data are available from the corresponding author upon reasonable request.

References

1. Dittmar T, Stubbins A. Dissolved Organic Matter in Aquatic Systems. *Treatise on Geochemistry, 2nd edn Elsevier: Oxford* 2014**:** 125-156.

2. Zhang C, Dang H, Azam F, Benner R, Legendre L, Passow U*, et al.* Evolving paradigms in biological carbon cycling in the ocean. *Natl Sci Rev* 2018, **5**(4)**:** 481-499.

3. Kujawinski EB. The Impact of Microbial Metabolism on Marine Dissolved Organic Matter. *Annu Rev Mar Sci* 2011, **3**(1)**:** 567-599.

4. Lennon J, Pfaff L. Source and supply of terrestrial organic matter affects aquatic microbial metabolism. *Aquat Microb Ecol* 2005, **39:** 107-119.

5. Berggren M, Ström L, Laudon H, Karlsson J, Jonsson A, Giesler R*, et al.* Lake secondary production fueled by rapid transfer of low molecular weight organic carbon from terrestrial sources to aquatic consumers. *Ecol Lett* 2010, **13**(7)**:** 870-880.

6. Cherif M, Loreau M. Stoichiometric Constraints on Resource Use, Competitive Interactions, and Elemental Cycling in Microbial Decomposers. *Am Nat* 2007, **169**(6)**:** 709-724.

7. Kellerman AM, Kothawala DN, Dittmar T, Tranvik LJ. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat Geosci* 2015, **8**(6)**:** 454-457.

8. Roth V-N, Lange M, Simon C, Hertkorn N, Bucher S, Goodall T*, et al.* Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat Geosci* 2019, **12**(9)**:** 755-761.

9. Judd KE, Crump BC, Kling GW. Variation in Dissolved Organic Matter Controls Bacterial Production and Community Composition. *Ecology* 2006, **87**(8)**:** 2068-2079.

10. Amaral V, Graeber D, Calliari D, Alonso C. Strong linkages between DOM optical properties and main clades of aquatic bacteria. *Limnol Oceanogr* 2016, **61**(3)**:** 906-918.

11. Logue JB, Stedmon CA, Kellerman AM, Nielsen NJ, Andersson AF, Laudon H*, et al.* Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. *ISME J* 2015, **10**(3)**:** 533-545.

12. Li X-M, Chen Q-L, He C, Shi Q, Chen S-C, Reid BJ*, et al.* Organic Carbon Amendments Affect the Chemodiversity of Soil Dissolved Organic Matter and Its Associations with Soil Microbial Communities. *Environ Sci Technol* 2018, **53**(1)**:** 50-59.

13. Orland C, Yakimovich KM, Mykytczuk NCS, Basiliko N, Tanentzap AJ. Think global, act local: The small‐scale environment mainly influences microbial community development and function in lake sediment. *Limnol Oceanogr* 2020, **65**(S1).

14. Osterholz H, Singer G, Wemheuer B, Daniel R, Simon M, Niggemann J*, et al.* Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *ISME J* 2016, **10**(7)**:** 1717-1730.

15. Tanentzap AJ, Fitch A, Orland C, Emilson EJS, Yakimovich KM, Osterholz H*, et al.* Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc Natl Acad Sci U S A* 2019, **116**(49)**:** 24689.

16. Danczak RE, Chu RK, Fansler SJ, Goldman AE, Graham EB, Tfaily MM*, et al.* Using metacommunity ecology to understand environmental metabolomes. *Nat Commun* 2020, **11**(1)**:** 6369.

17. Li H-Y, Wang H, Wang H-T, Xin P-Y, Xu X-H, Ma Y*, et al.* The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. *Microbiome* 2018, **6**(1).

18. Sarmento H, Gasol JM. Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ Microbiol* 2012, **14**(9)**:** 2348-2360.

19. Baines SB, Pace ML. The production of dissolved organic matter by phytoplankton and its importance to bacteria: Patterns across marine and freshwater systems. *Limnol Oceanogr* 1991, **36**(6)**:** 1078-1090.

20. Zhang Y, van Dijk MA, Liu M, Zhu G, Qin B. The contribution of phytoplankton degradation to chromophoric dissolved organic matter (CDOM) in eutrophic shallow lakes: field and experimental evidence. *Water Res* 2009, **43**(18)**:** 4685-4697.

21. Herzsprung P, Wentzky V, Kamjunke N, von Tümpling W, Wilske C, Friese K*, et al.* Improved Understanding of Dissolved Organic Matter Processing in Freshwater Using Complementary Experimental and Machine Learning Approaches. *Environ Sci Technol* 2020, **54**(21)**:** 13556-13565.

22. Sarmento H, Morana C, Gasol JM. Bacterioplankton niche partitioning in the use of phytoplankton-derived dissolved organic carbon: quantity is more important than quality. *ISME J* 2016, **10**(11)**:** 2582-2592.

23. Worm B, Lotze HK, Hillebrand H, Sommer U. Consumer versus resource control of species diversity and ecosystem functioning. *Nature* 2002, **417**(6891)**:** 848-851.

24. Chase JM. Stochastic Community Assembly Causes Higher Biodiversity in More Productive Environments. *Science* 2010, **328**(5984)**:** 1388.

25. Kassen R, Buckling A, Bell G, Rainey P. Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* 2000, **406:** 508-512.

26. Finke DL, Snyder WE. Niche Partitioning Increases Resource Exploitation by Diverse Communities. *Science* 2008, **321**(5895)**:** 1488.

27. Muscarella ME, Boot CM, Broeckling CD, Lennon JT. Resource heterogeneity structures aquatic bacterial communities. *ISME J* 2019, **13**(9)**:** 2183-2195.

28. Danczak RE, Goldman AE, Chu RK, Toyoda JG, Garayburu-Caruso VA, Tolić N*, et al.* Ecological theory applied to environmental metabolomes reveals compositional divergence despite conserved molecular properties. *bioRxiv* 2020**:** 2020.2002.2012.946459.

29. Fierer N, Bradford MA, Jackson RB. Toward an ecological classification of soil bacteria. *Ecology* 2007, **88**(6)**:** 1354-1364.

30. Dormann CF, Frund J, Bluthgen N, Gruber B. Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. *The Open Ecology Journal* 2009, **2:** 7-24.

31. Bluthgen N, Menzel F, Bluthgen N. Measuring specialization in species interaction networks. *BMC Ecol* 2006, **6:** 9.

32. Wang J, Pan F, Soininen J, Heino J, Shen J. Nutrient enrichment modifies temperature-biodiversity relationships in large-scale field experiments. *Nat Commun* 2016, **7:** 13960.

33. Muggeo V. Segmented: An R Package to Fit Regression Models With Broken-Line Relationships. *R News* 2008, **8:** 20-25.

34. Ellis N, Smith SJ, Pitcher CR. Gradient forests: calculating importance gradients on physical predictors. *Ecology* 2012, **93**(1)**:** 156-168.

35. Gower JC. Generalized procrustes analysis. *Psychometrika* 1975, **40**(1)**:** 33-51.

36. Hurley J, Cattell R. The PROCRUSTES program: producing direct rotation to test a hypothesized factor structure. *Behav Sci* 2007, **7:** 258-262.

37. Friedman J, Alm EJ. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput Biol* 2012, **8**(9)**:** e1002687.

38. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics* 1971, **27:** 857-871.

39. Malik AA, Martiny JBH, Brodie EL, Martiny AC, Treseder KK, Allison SD. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J* 2020, **14**(1)**:** 1-9.

40. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S*, et al.* The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* 2009, **106**(37)**:** 15527-15533.

41. Grace JB, Schoolmaster Jr. DR, Guntenspergen GR, Little AM, Mitchell BR, Miller KM*, et al.* Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere* 2012, **3**(8)**:** art73.

42. Breiman L. Random Forests. *Mach Learn* 2001, **45**(1)**:** 5-32.

43. Kothawala DN, Kellerman AM, Catalan N, Tranvik LJ. Organic Matter Degradation across Ecosystem Boundaries: The Need for a Unified Conceptualization. *Trends Ecol Evol* 2021, **36**(2)**:** 113-122.

44. Kellerman AM, Guillemette F, Podgorski DC, Aiken GR, Butler KD, Spencer RGM. Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. *Environ Sci Technol* 2018, **52**(5)**:** 2538-2548.

45. Davidson EA, Janssens IA. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* 2006, **440**(7081)**:** 165-173.

46. Sinha E, Michalak AM, Balaji V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* 2017, **357**(6349)**:** 405.

47. Anderson NJ, Bennion H, Lotter AF. Lake eutrophication and its implications for organic carbon sequestration in Europe. *Global Change Biol* 2014, **20**(9)**:** 2741-2751.

48. Wieder WR, Allison SD, Davidson EA, Georgiou K, Hararuk O, He Y*, et al.* Explicitly representing soil microbial processes in Earth system models. *Global Biogeochem Cy* 2015, **29**(10)**:** 1782-1800.

49. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK*, et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010, **7**(5)**:** 335-336.

50. Choi JH, Jang E, Yoon YJ, Park JY, Kim T-W, Becagli S*, et al.* Influence of Biogenic Organics on the Chemical Composition of Arctic Aerosols. *Global Biogeochem Cy* 2019, **33**(10)**:** 1238-1250.

51. 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* 2008, **6**(6)**:** 230-235.

52. Tolić N, Liu Y, Liyu A, Shen Y, Tfaily MM, Kujawinski EB*, et al.* Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra. *Anal Chem* 2017, **89**(23)**:** 12659-12665.

53. Kujawinski EB, Behn MD. Automated Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of Natural Organic Matter. *Anal Chem* 2006, **78**(13)**:** 4363-4373.

54. Koch BP, Dittmar T, Witt M, Kattner G. Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter. *Anal Chem* 2007, **79**(4)**:** 1758-1763.

55. Kim S, Kramer RW, Hatcher PG. Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Anal Chem* 2003, **75**(20)**:** 5336-5344.

56. Legendre P, Legendre L. *Numerical Ecology*. Elsevier Scientific: Oxford, U.K., 1998.

57. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003, **14:** 927-930.

58. Koch BP, Dittmar T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun Mass Spectrom* 2016, **30**(1)**:** 250-250.

59. LaRowe DE, Van Cappellen P. Degradation of natural organic matter: A thermodynamic analysis. *Geochim Cosmochim Acta* 2011, **75**(8)**:** 2030-2042.

60. Hughey CA, Hendrickson CL, Rodgers RP, Marshall AG, Qian K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal Chem* 2001, **73**(19)**:** 4676-4681.

61. Song H-S, Stegen JC, Graham EB, Lee J-Y, Garayburu-Caruso VA, Nelson WC*, et al.* Representing Organic Matter Thermodynamics in Biogeochemical Reactions via Substrate-Explicit Modeling. *Frontiers in Microbiology* 2020, **11**(2625).

62. Schneidman-Duhovny D, Bramer LM, White AM, Stratton KG, Thompson AM, Claborne D*, et al.* ftmsRanalysis: An R package for exploratory data analysis and interactive visualization of FT-MS data. *PLoS Comput Biol* 2020, **16**(3)**:** e1007654.

63. Lavorel S, Grigulis K, McIntyre S, Williams NSG, Garden D, Dorrough J*, et al.* Assessing functional diversity in the field – methodology matters! *Funct Ecol* 2007, **22:** 134-147.

64. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 2001, **129**(2)**:** 169-178.

65. Hiemstra PH, Pebesma EJ, Twenhöfel CJW, Heuvelink GBM. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput Geosci* 2009, **35**(8)**:** 1711-1721.

66. Dormann CF, Strauss R, Peres-Neto P. A method for detecting modules in quantitative bipartite networks. *Methods Ecol Evol* 2014, **5**(1)**:** 90-98.

67. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics* 2014, **30**(19)**:** 2811-2812.

68. Liaw A, Wiener M. Classification and Regression by randomForest. *R News* 2002, **2:** 18-22.

69. Borcard D, Legendre P, Drapeau P. Partialling out the spatial component of ecological variation. *Ecology* 1992, **73:** 1045–1055.

70. Miller JK, Farr SD. Bimultivariate redundancy: a comprehensive measure of interbattery relationship. *Multivar Behav Res* 1971, **6:** 313–324.

71. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara RB*, et al.* vegan: Community Ecology Package. CRAN R package. 2017.

72. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med* 2016, **4**(7)**:** 136.

73. Sakamoto Y, Ishiguro M, Kitagawa G. *Akaike Information Criterion Statistics*. D. Reidel Publishing Company: Boston, 1986.

74. Hu A, Wang J, Sun H, Niu B, Si G, Wang J*, et al.* Mountain biodiversity and ecosystem functions: interplay between geology and contemporary environments. *ISME J* 2020, **14**(4)**:** 931-944.

75. Ishwaran H, Kogalur UB. Random survival forests for R. *R News* 2007, **7:** 25-31.

76. Ishwaran H, Kogalur UB. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). 2019.

77. Grace JB, Anderson TM, Seabloom EW, Borer ET, Adler PB, Harpole WS*, et al.* Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* 2016, **529:** 390.

78. Grace JB, Anderson TM, Olff H, Scheiner SM. On the specification of structural equation models for ecological systems. *Ecol Monogr* 2010, **80**(1)**:** 67-87.

79. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw* 2012, **48**(2)**:** 1-36.

Material and 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*. (2016) [32]. The annual temperatures ranged from 4.2-12.9 °C in China and -2.9-0.7 °C in Norway. 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 3,822, 3,505, 2,915, 2,580 and 2,286 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, which included 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^{-1} of KNO_3. 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 as previously described in Wang *et al*. (2016) [32]. To compensate for nitrate additions shifting stoichiometric ratios, KH_2PO_4 was added to 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 (14.49). 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^{-1}; Fig. S27). 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, and the sediment bacteria were examined using high-throughput sequencing of 16S rRNA genes. The

676 sequences were processed in QIIME (v1.9) [49] and OTUs were defined at 97% sequence
677 similarity. The bacterial sequences were rarefied to 20,000 per sample. Further details on
678 field experiments, sample collection, physicochemical and bacterial community analyses
679 are available in Wang *et al.* (2016).

680

681 **ESI FT-ICR MS analysis of DOM samples**

682 Highly accurate mass measurements of DOM within the sediment samples were
683 conducted using a 15 Tesla solariX XR system, a ultrahigh-resolution Fourier transform
684 ion cyclotron resonance mass spectrometer (FT-ICR MS, Bruker Daltonics, Billerica,
685 MA) coupled with an electrospray ionization (ESI) interface, as demonstrated previously
686 [50] with some modifications. DOM was solid-phase extracted (SPE) with Agilent VacElut
687 resins before FT-ICR MS measurement [51] with minor modifications. Briefly, an aliquot
688 of 0.7 g freeze-dried sediment was sonicated with 30 ml ultrapure water for 2 h, and
689 centrifuged at 5,000 g for 20 min. The extracted water was filtered through the 0.45 μm
690 Millipore filter and further acidified to pH 2 using 1 M HCl. Cartridges were drained,
691 rinsed with ultrapure water and methanol (ULC-MS grade), and conditioned with pH 2
692 ultrapure water. Calculated volumes of extracts were slowly passed through cartridges
693 based on DOC concentration. Cartridges were rinsed with pH 2 ultrapure water and dried
694 with N_2 gas. Samples were finally eluted with methanol into precombusted amber glass
695 vials, dried with N_2 gas and stored at -20 °C until DOM analysis. The extracts were
696 continuously injected into the standard ESI source with a flow rate of 2 μl min^{-1} and an
697 ESI capillary voltage of 3.5 kV in negative ion mode. One hundred single scans with a
698 transient size of 4 mega words, an ion accumulation time of 0.3 s, and within the mass
699 range of m/z 150-1200, were co-added to a spectrum with absorption mode for phase
700 correction, thereby resulting in a resolving power of 750,000 (FWHM at m/z 400). All
701 FT-ICR mass spectra were internally calibrated using organic matter homologous series
702 separated by 14 Da (-CH_2 groups). The mass measurement accuracy was typically within
703 1 ppm for singly charged ions across a broad m/z range (150-1,200 m/z).

704 Data Analysis software (BrukerDaltonik version 4.2) was used to convert raw
705 spectra to a list of m/z values using FT-MS peak picker with a signal-to-noise ratio (S/N)

706 threshold set to 7 and absolute intensity threshold to the default value of 100. Putative
707 chemical formulae were assigned using the software Formularity [52] following the
708 Compound Identification Algorithm [53]. In total, 19,538 molecular formulas were
709 putatively assigned for all samples (n = 300) based on the following criteria: S/N > 7, and
710 mass measurement error < 1 ppm, considering the presences of C, H, O, N, S and P and
711 excluding other elements or an isotopic signature. All formula assignments were further
712 screened to meet the criteria as follows [54]: (1) formulae containing an odd number of
713 nitrogen atoms had an even nominal m/z and those containing an even number of
714 nitrogen atoms had an odd nominal m/z; (2) the number of hydrogen atoms was at least
715 1/3 of carbon and could not exceed 2C+N+2; (3) the number of nitrogen or oxygen atoms
716 could not exceed the number of carbon atoms; (4) the ratio of O/C was set to 0-1, H/C ≥
717 0.3, N/C ≤ 1, double bond equivalents (DBE) ≥ 0.

718 The assigned molecules were categorised into eight compound classes or 12
719 elemental combinations. The compound classes based on van Krevelen diagrams [55] were
720 lipids (O/C = 0-0.3, H/C = 1.5-2.0), proteins and amino sugars (O/C = 0.3-0.67, H/C =
721 1.5-2.2), carbohydrates (Carb; O/C = 0.67-1.2, H/C = 1.5-2), unsaturated hydrocarbons
722 (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
723 (O/C = 0.67-1.2, H/C = 0.5-1.5), and condensed aromatics (ConHC; O/C = 0-0.67, H/C =
724 0.2-0.7). The elemental combinations were CH, CHN, CHNO, CHNOP, CHNOS,
725 CHNOSP, CHNS, CHO, CHOP, CHOS, CHOSP and CHS.

726

727 **Estimating DOM features**

728 We considered DOM features from three aspects: alpha diversity, beta diversity
729 and molecular traits. These features were considered for all molecules (19,538 different
730 formulae), but also for subsets of molecules within each category of compound classes or
731 elemental combinations. The dataset based on all molecular formulae was named "All
732 molecules", while the datasets of subsets of formulae were named by "category name +
733 compounds". The relative abundance of molecules was calculated by normalizing signal
734 intensities of assigned peaks to the sum of all intensities within each sample. We
735 considered two additional 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) [56]. The analyses of chemical diversity were performed using the R package vegan V2.4.6 [57].

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}) [58], DBE [58], DBE minus oxygen (DBE_O) [58], DBE minus AI (DBE_{AI}) [58], standard Gibb's Free Energy of carbon oxidation (GFE) [59], Kendrick Defect ($kdefect_{CH2}$) [60], 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}) [61]. All calculations were performed using the R package ftmsRanalysis V1.0.0 [62] and the scripts at https://github.com/danczakre/ICRTutorial. DBE represents the number of unsaturated bonds and rings in a molecule [58]. 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 [61]. 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 ($Mass_{wm} = \Sigma(Mass_i \times I_i) / \Sigma(I_i)$) using the R package FD V1.0.12 [63]. In addition, ten molecular sub-mixtures 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.

Estimating bacterial communities

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

771

772 **Estimating associations between DOM and microbes**

773 At the DOM composition level, we examined DOM-microbe associations from
774 the following aspects: Pearson's correlation between alpha diversity of DOM and bacteria,
775 and a Mantel correlation between the beta diversity of DOM and bacteria. We also tested
776 the congruence between DOM and bacterial composition using Procrustes analysis of
777 NMDS coordinates estimated for each community across elevations and nutrient
778 enrichment levels with the Bray-Curtis dissimilarity metric [35, 36]. Procrustes analysis is a
779 technique for comparing the relative positions of points in two multivariate datasets. It
780 attempts to stretch and rotate the points in one matrix, such as points obtained from a
781 NMDS, to be as close as possible to points in another matrix, thus preserving the relative
782 distances between points within each matrix [35, 36]. This procedure yields a measure of fit,
783 M^2, which is the sum of squared distances between corresponding data points after the
784 transformation. Pointwise residuals indicate the difference between two different
785 community ordinations for each sample. The statistical significance of the Procrustes
786 analysis (i.e., M^2) can then be assessed by randomly permutating the data 1,000 times [64].
787 This analysis was performed using the R package vegan V2.4.6.

788 We further quantified DOM-microbe associations at a molecular level using two
789 different co-occurrence analyses. First, Spearman's rank correlation coefficient ρ was
790 calculated between the relative abundance of each molecule m/z ion and bacterial OTU
791 (or genus). For each molecule, we then calculated the Spearman ρ difference by
792 subtracting the mean absolute ρ value of the negative correlations across all bacterial
793 OTUs from the mean of the positive correlations. Larger positive and negative values
794 indicate that molecules were more strongly positively and negatively correlated with

795 bacterial communities, respectively. The relationships among the Spearman ρ difference,
796 H/C and O/C were summarised using kriging interpolation with the R package automap
797 V1.0.14 [65]. Second, SparCC (Sparse Correlations for Compositional data) was applied to
798 build DOM-microbe bipartite networks. SparCC is a correlation method that can infer the
799 interrelationships between DOM and bacteria for compositional data with higher
800 accuracy [37] than general correlation approaches, such as Spearman's correlation, because
801 it explicitly assumes that the underlying networks have many missing associations. We
802 used bacterial genera rather than OTUs for bipartite network analysis because there were
803 over 20,000 and 10,000 bacterial OTUs for Norway and China, respectively, and there
804 are computational limits on handling such large bipartite networks for the analyses
805 described in the next paragraph. However, using bacterial genera was reasonable as
806 individual DOM-bacteria associations were similar for both bacterial OTUs and genera
807 ($R^2 > 0.80$, $P \leq 0.001$; Fig. S14). Similar conclusions were also obtained with either
808 OTUs or genera when relating the pairwise distances of molecular traits with SparCC ρ
809 values among DOM molecules in Fig. 3c. To reduce type I errors in the correlation
810 calculations created by low-occurrence genera or molecules, the majority rule was
811 applied, retaining genera or molecules observed in more than half of the total samples ($\geq$
812 75 samples) in China or Norway. The filtered table, including 1,340 and 1,246 DOM
813 molecules, and 75 and 49 bacterial genera in China and Norway, respectively, was then
814 used for pairwise correlation calculation of DOM and bacteria using SparCC with default
815 parameters [37].

816 Finally, bipartite network analysis at a molecule and network level was performed
817 to quantify the specialization of DOM-microbe associations. The threshold correlation for
818 inclusion in bipartite networks was $|\rho| = 0.30$ to exclude weak interactions and we
819 retained the adjacent matrix with only the interactions between DOM and bacteria. We
820 then constructed two types of networks based on negative and positive correlations
821 (SparCC $\rho \leq -0.30$ and $\rho \geq 0.30$, respectively). The SparCC ρ values were multiplied by
822 10,000 and rounded to integers, and the absolute values were taken for negative networks
823 to enable the calculations of specialization indices. A separate negative and positive
824 network was obtained for each microcosm based on its species composition. For the

network level analysis, we calculated H_2', a measure of specialization [30], for each network:

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

$$H_2' = \frac{H_{2max}-H_2}{H_{2max}-H_{2min}},$$

where $p_{ij} = a_{ij}/m$, represents the proportion of interactions in a i × j matrix. a_{ij} is number of interactions between DOM molecule i and bacterial genus j, which is also referred as "link weight". m is 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:

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

$$d_i' = \frac{d_i-d_{min}}{d_{max}-d_{min}},$$

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.

Both specialization indices consider interaction abundance and are standardised to account for heterogeneity in the interaction strength and species richness. 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 $d' = \Sigma(d'_i \times I_i) / \Sigma(I_i)$. 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) [31] (Fig. S28). 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 ($S_{observed}$; that is, H_2', d' of DOM, and d' of bacteria) by calculating their z-scores [66] using the equation $z_S = (S_{observed} - \overline{S}_{null}) / (\sigma_{S_{null}})$ where $\overline{S}_{null}$ and $\sigma_{S_{null}}$ were, respectively, the mean and standard deviation of the null distribution of S (S_{null}). One hundred randomised null networks were generated for each bipartite network to derive S_{null} using the *swap.web* algorithm, which keeps species richness and the number of interactions per species constant along with network connectance. The relationships among H_2', weighted means of d' for DOM molecules and bacterial genera were compared using kriging interpolation with the R package automap V1.0.14. The obtained network was visualised using circlize V0.4.10 [67] and analysed using the R package bipartite V2.15 [30].

Statistical analyses

We used the following explanatory variables related to distal and proximal controls on DOM-microbe associations. Distal environmental drivers included climate change (i.e., water temperature), human impacts (i.e., nutrient enrichment), and contemporary nutrients (i.e., sediment total nitrogen (TN), total phosphorus (TP), NO_x^-, NO_2^-, NH_4^+ and PO_4^{3-}, and water NO_3^-, NO_2^-, NH_4^+ and PO_4^{3-}). Proximal drivers included energy supply (i.e., sediment total organic carbon, dissolved organic carbon, water pH and sediment Chlorophyll *a* (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, AI_{Mod}, DBE, DBE_O, DBE_{AI}, GFE, $kdefect_{CH2}$, NOSC, O/C, H/C, N/C, P/C, S/C and Y_{met}). Detailed information about these explanatory variables is listed in Table S1. 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 [32]. The response variables included DOM features (i.e., alpha diversity, beta diversity and molecular traits)

and DOM-microbe network statistics (e.g., H_2'), and were analysed for their patterns and underlying drivers along the two main environmental gradients: elevation and nutrient enrichment.

(1) Patterns of DOM features and DOM-microbe associations along the environmental gradients

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 [33]. These breakpoint estimations were supported by gradient forest analysis [34], which was used to assess the DOM compositional changes and important breakpoints across multiple molecules along the gradient of nutrient enrichment. This analysis produces the standardised density of splits, that is the kernel density of splits divided by the observation density, which shows where important changes in the abundance of multiple molecules occur along the nutrient gradient and indicates the compositional rate of change. 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 [34] and extendedForest V1.6.1 [68].

For DOM-microbe 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.

(2) Drivers of DOM features and DOM-microbe associations

To evaluate the key drivers of DOM features and DOM-microbe associations, we used variation partitioning analysis (VPA) [69], multiple regression, random forest analysis [42] and structural equation modelling (SEM) [41]. In particular, the first analysis disentangled the important roles of microbes from other explanatory variables, while the

latter three 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 S1). We selected explanatory variables for regression analyses by forward selection with Akaike information criterion (AIC) [70]. 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 [71].

Second, stepwise multiple regression was performed to test the statistical significance and predictive power of the net effects of diversity (i.e., biodiversity and chemodiversity) or molecular traits on the bipartite network specialization H_2'. The net effects of diversity or molecular traits were evaluated by the improvements in the explained variances relative to models without diversity and molecular traits (i.e., in models using only the variables associated with environments and energy supply). The analysis was conducted with forward selection of explanatory variables [72]. We chose the final model that had the lowest AIC value [73]. ANOVA was used to test the statistical significance of two models including or excluding diversity or molecular traits as predictors, and the increase in the model R^2 was determined as the net effects of diversity or molecular traits on H_2'.

Third, 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 (2,000 trees). More details on this method were described in previous literature [74]. In addition, random forest analysis was also used to test the net

effects of diversity or molecular traits on H_2'. This analysis was conducted using the R package randomForestSRC V2.8.0 [75, 76].

Finally, 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. S1. 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 [77], 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 S2). Based on all the hypothesised links among composite variables (that is, full model; Fig. S1), we examined all alternative models using AIC and overall model fit statistics [78]. We chose the final model to report as that with the lowest AIC value from models with a non-significant χ^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 S3). We implemented the SEMs using R package lavaan V.0.5.23 [79].

Predictions of DOM-microbe associations in Taihu Lake

Using the parameter estimates obtained from SEM fitted to the bipartite networks in subtropical China, we estimated spatiotemporal variation of DOM-microbe associations 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_{nut}), energy supply (P_{energy}), biodiversity (P_{biodiv}), chemodiversity ($P_{chemodiv}$) and molecular traits (P_{trait}) based on climate and eutrophication drivers:

969 $$P_{nut} = \lambda_{nut,temp} \times X_T + \lambda_{nut,N} \times X_N$$

970 $$P_{energy} = \lambda_{energy,temp} \times X_T + \lambda_{energy,N} \times X_N + \lambda_{energy,nut} \times P_{nut}$$

971 $$P_{biodiv} = \lambda_{biodiv,temp} \times X_T + \lambda_{biodiv,N} \times X_N + \lambda_{biodiv,nut} \times P_{nut} + \lambda_{biodiv,energy} \times P_{energy}$$

972 $$P_{chemodiv} = \lambda_{chemodiv,temp} \times X_T + \lambda_{chemodiv,N} \times X_N + \lambda_{chemodiv,nut} \times P_{nut} + \lambda_{chemodiv,energy} \times$$

973 P_{energy}

974 $$P_{trait} = \lambda_{trait,temp} \times X_T + \lambda_{trait,N} \times X_N + \lambda_{trait,nut} \times P_{nut} + \lambda_{trait,energy} \times P_{energy} + \lambda_{trait,biodiv} \times$$

975 $P_{biodiv} + \lambda_{trait,chemodiv} \times P_{chemodiv}$

976 where X_T and X_N were water temperature and total nitrogen, respectively, for the
977 32 sites across the whole Taihu Lake (Fig. S27a). The abbreviations of path coefficients
978 (λ) are detailed in Table S4.

979 Similarly, we calculated the specialization of DOM-microbe associations (Y_{H2})
980 using a linear equation: $Y_{H2} = \lambda_{H2,temp} \times X_T + \lambda_{H2,N} \times X_N + \lambda_{H2,nut} \times P_{nut} + \lambda_{H2,energy} \times P_{energy}$
981 $+ \lambda_{H2,biodiv} \times P_{biodiv} + \lambda_{H2,chemodiv} \times P_{chemodiv} + \lambda_{H2,trait} \times P_{trait}$. We used the predicted values
982 for contemporary nutrients, energy supply, biodiversity, chemodiversity and molecular
983 traits in the overall prediction model to account for the indirect effects of water
984 temperature and total nitrogen on specialization. The models were calculated with a
985 yearly time step based on the annual means of water temperature and total nitrogen for
986 each site during 2007-2018. The temporal changes in specialization were calculated using
987 2007 as a baseline to which all predictions were compared.

988 The above predictions aimed to apply our EDTiA framework to estimate changes
989 in DOM-microbe associations under temperature change and eutrophication in Taihu
990 Lake, and potential uncertainties in the estimated associations should however be noted
991 as follows. First, local environmental variation (e.g., N/P ratio changes) and different
992 microbial species pools between our field microcosms and natural lake sediments would
993 likely influence the accuracy of predictions. Second, spatial and temporal heterogeneity
994 of sediments would influence local environments and the composition of both DOM and
995 microbes and thus the projection of estimates across Taihu Lake. Third, the transferability
996 and extrapolation of SEM models to Taihu Lake would be one of the difficulties in

997 prediction practices. We thus selected the SEM models in China rather than Norway for
998 more similar climatic conditions to the target lake. The annual mean water temperatures
999 in Taihu Lake were covered by the temperature variations across the elevations between
1000 2,286 and 3,822 m a.s.l. in Laojun Mountain, and the annual mean total nitrogen fell into
1001 the gradient of nutrient concentrations between 0 and 36 mg N L^{-1}. Finally, lake
1002 management such as mechanical removal of algae would affect energy supply and
1003 consequently prediction accuracy.

1004

1005

Figure legends

Figure 1. A framework for studying the effects of global change on DOM-microbe associations. (a) DOM-microbe associations affected by the three proximal controls, namely energy supply and both the diversity and traits of DOM and microbes. The relationships among the three controls 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_1-C_i) and microbial species (circles M_1-M_j). 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, such as *d*' for DOM and microbes, which describes the levels of "vulnerability" of DOM molecules and "generality" of microbial species, respectively. (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 color 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 setup 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 [32].

Figure 2. DOM features and their microbial 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) and Norway (blue lines). 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. For better visualization, we did not include the data points in Fig. 2a but showed detailed scatter plots in Fig. S5. To visualise the compositional turnover of DOM, we plotted the standardised density of splits showing where important changes in the abundance of multiple molecules occurred along the nutrient gradient. The standardised density of splits was determined by gradient forest analysis [34]. (b) The congruence between DOM and bacterial compositions across different elevations in China and Norway was examined using Procrustes analysis [35, 36]. Each line with circle and triangle ends connects to a single community of DOM and bacteria, respectively, and is colored by elevation in either China (red) or Norway (blue). The fit of overall Procrustes transformation is reported as the M^2 value. (c) The effects of nutrient enrichment on DOM-microbe associations. The associations were quantified by the Pearson correlation coefficient r between alpha diversity of DOM and bacteria (upper panel), and by the Mantel r between the beta diversity of DOM and bacteria (lower panel). We then visualised these associations with loess regression models along the nutrient gradient. The colours of the lines indicate the DOM composition for all formulae and categories of compound classes or elemental combinations.

Figure 3. 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, we subtracted the mean absolute Spearman's rank correlation coefficient ρ 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 and positive bipartite networks between DOM molecules and bacterial genera in China or Norway estimated using SparCC (Sparse Correlations for Compositional data) [37]. 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. S17 and Table S1. A line connecting two nodes indicates an interaction between a DOM molecule and bacterial genus. (c-d) We examined the relationships

between molecular traits and the negative (left panel) or positive (right panel) DOM-microbe bipartite networks in China (upper panel) or Norway (lower panel). For all pairs of DOM molecules, we separately calculated pairwise Gower distances between the molecular traits and their SparCC ρ values with bacterial OTUs. Statistical significance between distance matrices was determined with a 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).

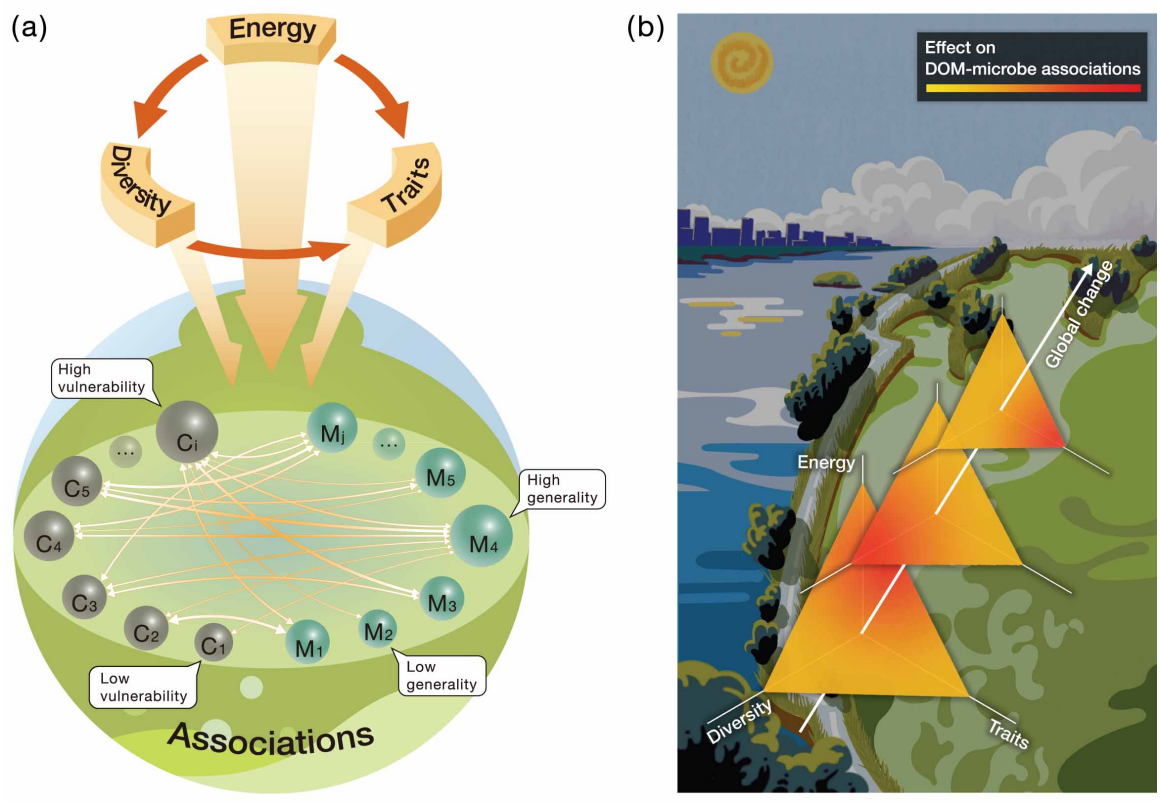
Figure 4. Relative importance of diversity and molecular traits in explaining specialization of DOM-microbe networks. (a) We plotted specialization H_2' against nutrient enrichment for negative (left panel) and positive (right panel) bipartite networks for each elevation in China (red lines) and Norway (blue lines). Statistical significance of linear model fits was indicated by solid ($P \leq 0.05$) or dotted ($P > 0.05$) lines. For better visualization, we omitted the data points but these are shown in Fig. S20. (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 colors of ovals or rectangles as in Fig. S1. Abbreviations of explanatory variables are detailed in Table S1.

Figure 5. Structural equation models [41] to explain specialization of DOM-microbe networks. 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 (a) the total and indirect effects of global change and human impacts via proximal variables and (b) the total and direct effects of proximal variables. Proximal variables

were energy supply, biodiversity, chemodiversity and molecular traits, and are described in detail in Table S2. Details of the full structural equation models are shown in Fig. S24.

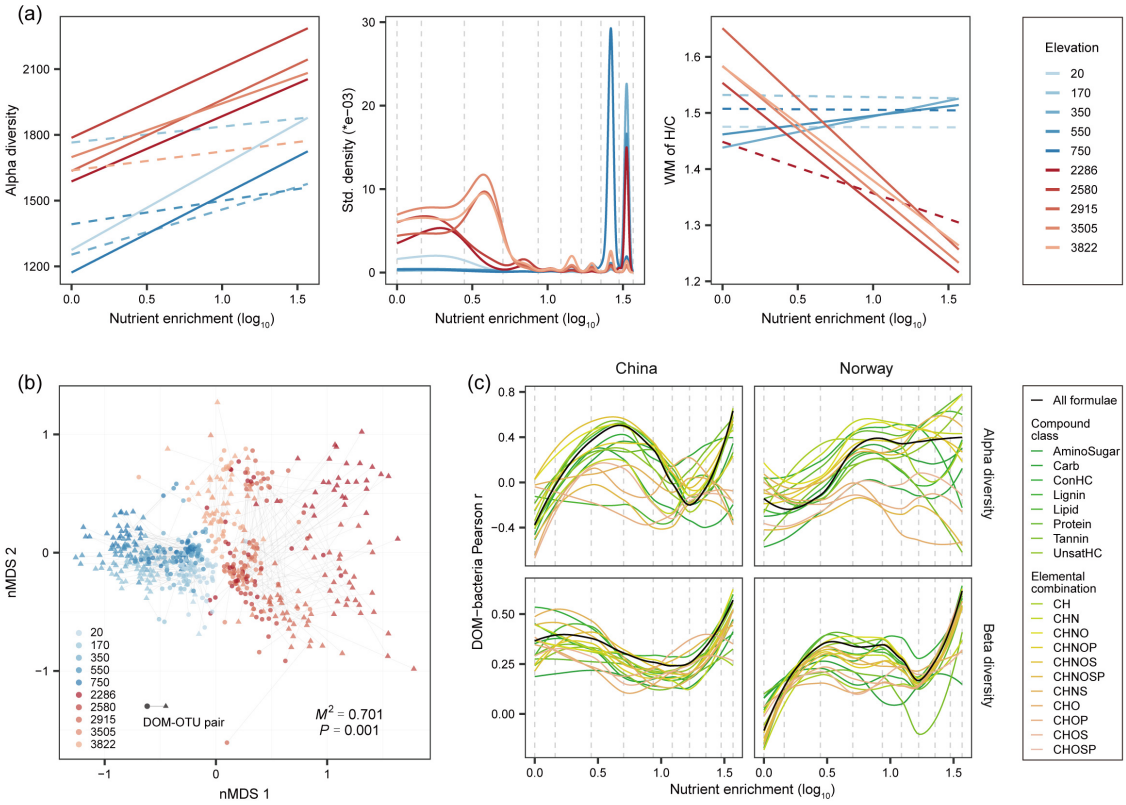
Figure 6. Decadal change in predicted specialization of DOM-microbe networks in Taihu Lake. (a) Changes in H_2' of negative (upper panel) and positive (lower panel) bipartite networks from 2007 to 2018. (b) The spatial distribution of changes in H_2' of negative (upper panel) and positive (lower panel) networks in 2018 across the Taihu Lake. Estimated changes in H_2' were calculated for the 32 sites across the whole of Taihu Lake (Fig. S27a) 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.

1111 **Figure 1**



1112

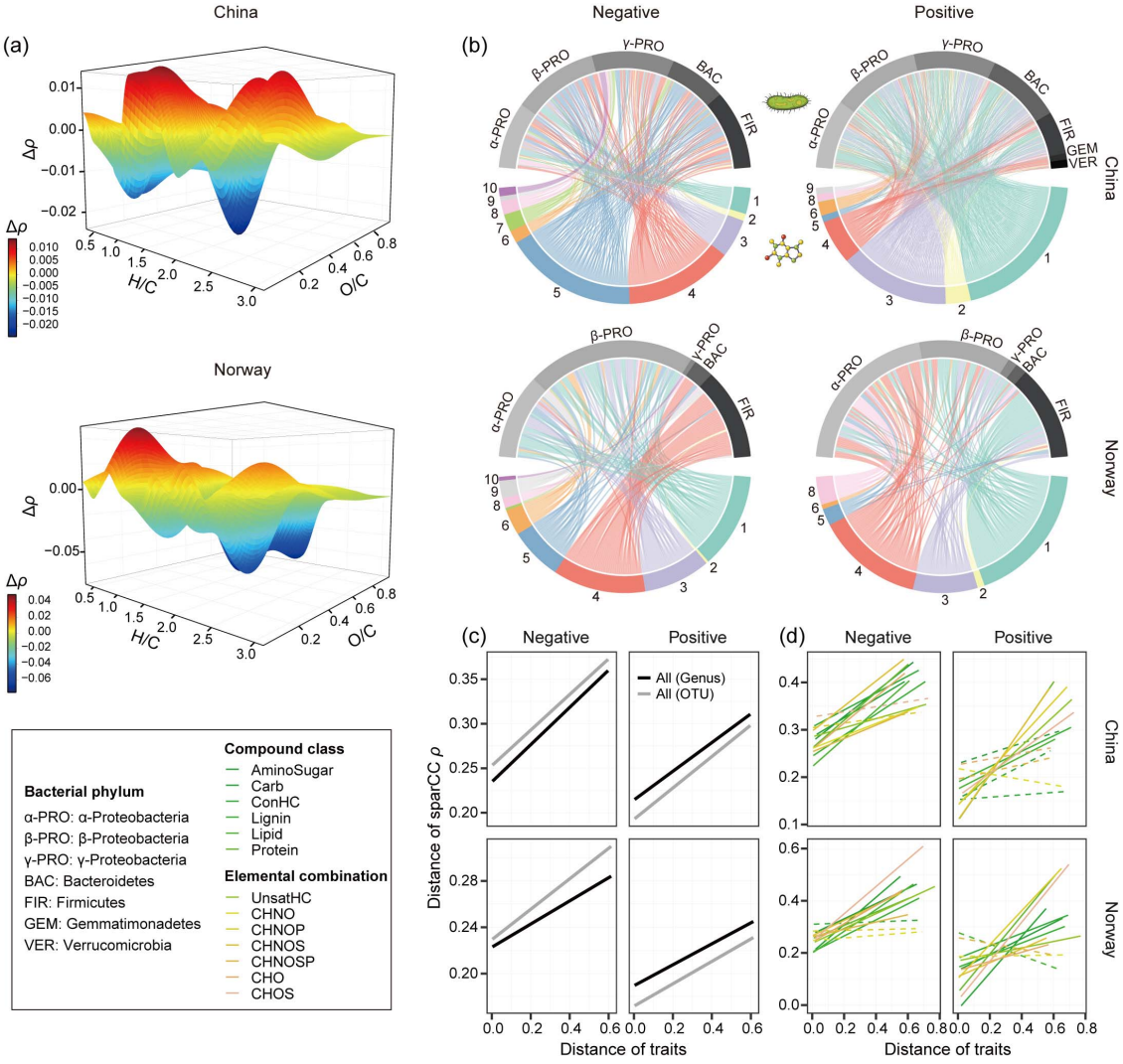
1113 **Figure 2**



1114

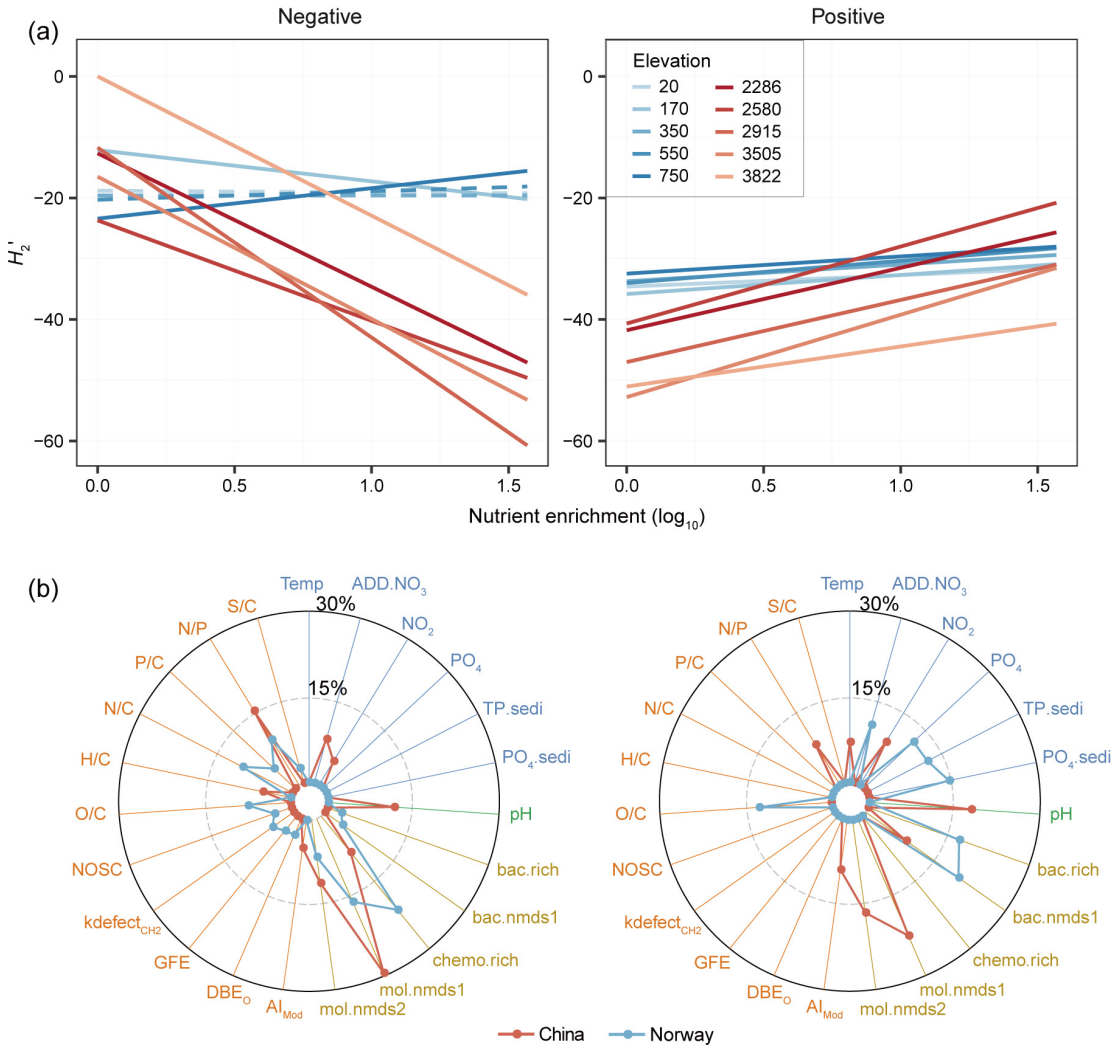
1115

Figure 3



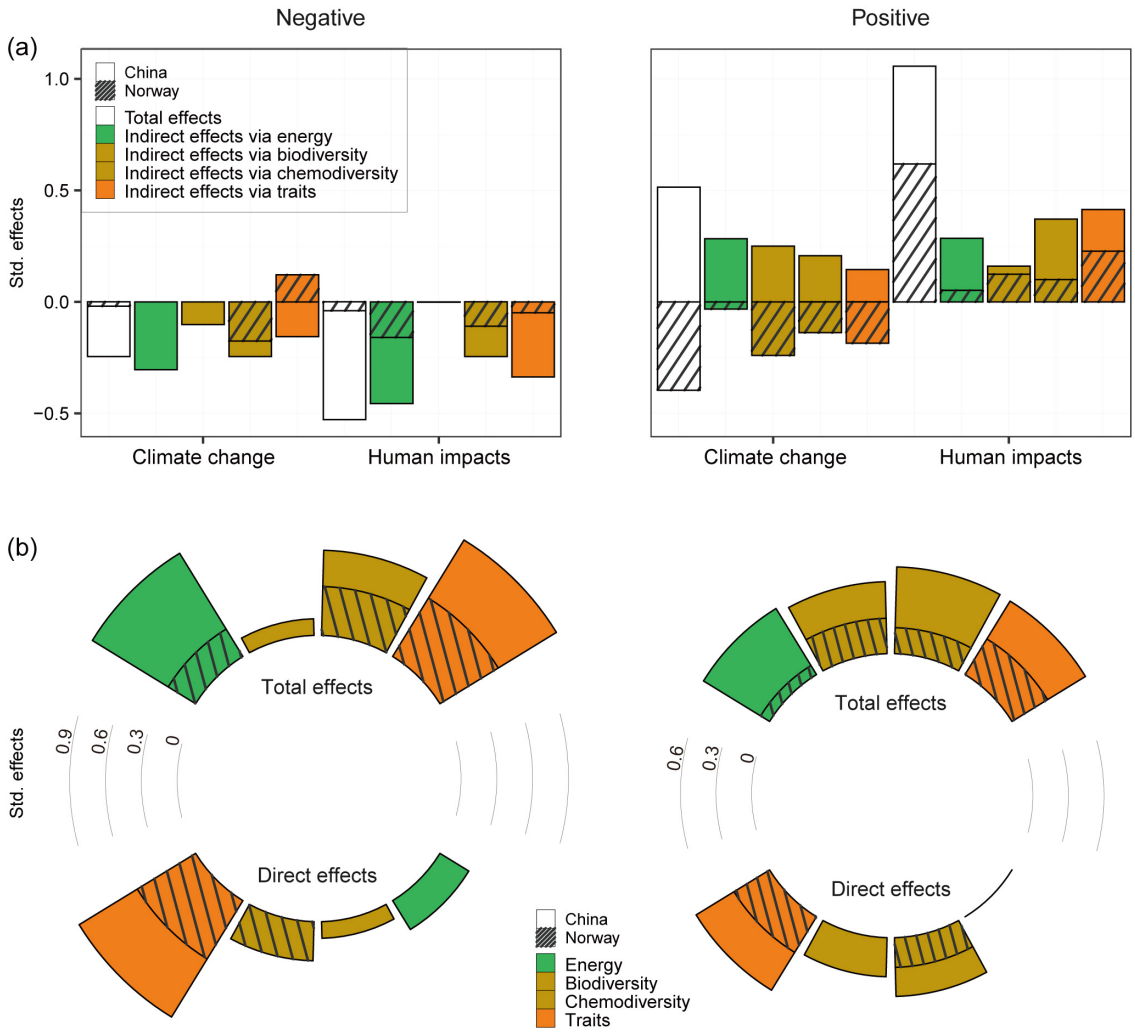
1116

1117 **Figure 4**



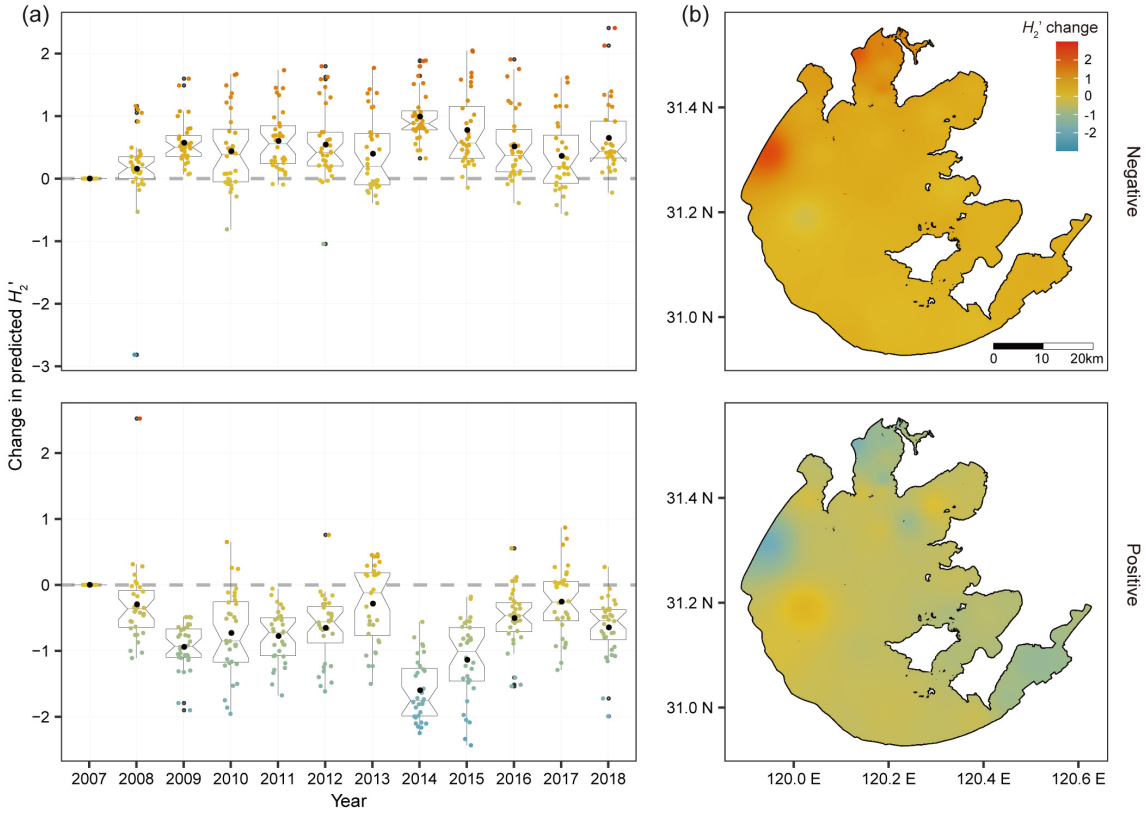
1118

1119 **Figure 5**



1120

1121 **Figure 6**



1122