Chemical diversity promotes ecosystem function

Jeremy A. Fonvielle[1,2*], Sarah R. Sandor[1], Thorsten Dittmar[3,4], Andrew J. Tanentzap[1,3,5*]

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

[2]Department of Biochemistry, Hopkins Building, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, United Kingdom

[3]Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University, Oldenburg 26129, Germany

[4]Helmholtz Institute for Functional Marine Biodiversity, Carl von Ossietzky University, Oldenburg 26129, Germany

[5]Ecosystems and Global Change Group, School of the Environment, Trent University, Peterborough, Ontario K9L 0G2, Canada

*Authors for correspondence: jaf91@cam.ac.uk; atanentzap@trentu.ca

Abstract

Nature benefits people in many ways that rely on how microorganisms and abiotic processes interact with organic matter to influence ecosystem function (EF). Despite many studies evaluating how biodiversity affects EF, most variation in EF remains unexplained. Here, we tested how the variety of organic compounds in the environment, termed chemodiversity, influences EF. Using a laboratory experiment and 101-lake survey spanning Europe, we discovered that chemodiversity, not biodiversity, consistently explained variation in multiple functions related to carbon cycling and ecosystem metabolism. We attributed positive chemodiversity-ecosystem multifunctionality relationships to the accumulation of compounds that could be biochemically transformed in more diverse ways at higher chemodiversity. We estimated chemodiversity will decline by 2100 due to changes in organic matter sources, causing European lakes to emit, on average, 1.6-times more carbon. These findings reframe our understanding of ecosystem functioning and suggest chemodiversity is key to sustain the benefits that nature delivers to people.

Introduction

Nature benefits society such as by regulating the climate and providing food and clean water[1,2]. These services are in part delivered by the multiple functions undertaken by microorganisms that break down and transform different organic substrates[3]. For example, as microbes assimilate detritus and increase in biomass[4], they support the growth of higher trophic levels[5], as well as modify organic compounds into forms that sequester carbon longer in the environment and thereby contribute to climate regulation[6]. Consequently, ecosystem function is enhanced by both a greater diversity of metabolic capabilities in microbial communities[3] and the environmental conditions that favour this biodiversity[7]. However, large amounts of variation in the capacity of ecosystems to provide multiple functions simultaneously, hereafter termed ecosystem multifunctionality[1], remain unexplained. Many large-scale studies have found that biodiversity only explains between 2 to 20% of the variation in ecosystem multifunctionality[3,8,9], though there are some exceptions[10].

The organic substrates that microbes use to influence biological, geochemical, and physical processes within ecosystems are an overlooked source of variation in past studies that predict ecosystem multifunctionality[6,11,12]. Natural organic matter in aquatic and terrestrial ecosystems contains tens of thousands (and likely millions) of unique molecules of varying origin and composition, making it one of the most complex mixtures on Earth[13]. This variety of molecules within organic matter and their characteristics, such as their bioavailability and stoichiometry, are together referred to as chemodiversity[14]. Chemodiversity can be described by measuring the diversity and abundance of molecular formulae and their characteristics, similar to approaches for studying biological mixtures of species[15]. Despite many surveys of chemodiversity in waters and soils[16,17], there has been no attempt to ask how chemodiversity

53 influences ecosystem function in nature as compared with biodiversity and other environmental
54 variables.
55
56 We hypothesised that chemodiversity influences ecosystem multifunctionality in two main
57 ways. First, the chemical properties of organic compounds (e.g. structure, elemental
58 composition, and biological and photochemical reactivity) determine the biotic and abiotic
59 pathways by which they are degraded and transformed[18,19]. A greater number of compounds
60 can therefore directly promote ecosystem multifunctionality by increasing the number of ways
61 that microbial communities and abiotic processes transform organic matter[13]. For example,
62 adding compounds to a mixture can increase the variety of unique chemical properties and thus
63 the number of organic compounds that are degraded, subsequently increasing microbial
64 biomass. Second, chemodiversity can indirectly influence ecosystem multifunctionality by
65 changing the physical environment. For example, because compounds vary in their ability to
66 absorb light[20], the addition of light-absorbing compounds to aquatic ecosystems can reduce
67 photosynthetic rates and limit heat penetration into deeper waters[21,22]. These changes can result
68 in warmer and more nutrient-rich surface waters that accelerate microbial activity and carbon
69 cycling between the land surface and atmosphere[6,23]. For these reasons, the number and
70 characteristics of organic compounds in the environment can provide an integrative measure
71 across multiple ecosystem functions whose importance can be compared directly with other
72 variables traditionally used to explain ecosystem multifunctionality, including biodiversity[9]
73 and climate[3,8]. However, whether chemodiversity explains a relatively large amount of
74 variation in ecosystem multifunctionality remains untested.
75
76 Here we tested the hypothesis that chemodiversity promotes ecosystem multifunctionality by
77 combining a manipulative lab experiment with the largest field survey of these variables to

78 date. We focused on lakes unlike past studies that tested associations between biodiversity and
79 ecosystem multifunctionality in terrestrial ecosystems[8]. Despite covering only 3.7% of the
80 Earth's surface, the functions of lakes, including carbon cycling and food web production[23,24],
81 are disproportionately important to many services valued by society, such as climate regulation
82 and food provisioning. We first used a lab experiment to test how changing chemodiversity
83 changes ecosystem multifunctionality, and then tested the importance of this association in
84 natural lakes that span 30° of latitude as compared with known drivers of ecosystem
85 multifunctionality, namely biodiversity and physicochemical variables. We defined
86 multifunctionality from a human perspective that values services like climate regulation and
87 food web productivity, and measured chemodiversity in surface waters using ultrahigh-
88 resolution Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to
89 identify the diversity of compounds involved in biotic and abiotic reactions. Although the
90 analytical window of FT-ICR MS can omit low (<150 m/z) and high-molecular-mass (>1,000
91 m/z) compounds that contribute to biotic and abiotic reactions[25], it is the most sensitive tool
92 available for characterising chemodiversity and captures most of the organic compounds that
93 interact with microbes[26]. We then used the field survey and current theories that synthesise how
94 different types of chemodiversity originate[13,27] to develop a general understanding of how and
95 where chemodiversity matters for ecosystem multifunctionality. As lakes are more abundant at
96 northern latitudes[28] that are undergoing large changes in organic matter inputs, we use our
97 results to forecast the consequences of environmental change for chemodiversity and the
98 potential impacts on the benefits that lakes provide to society.

99

100 *Chemodiversity is the main driver of ecosystem multifunctionality*

101 To test if ecosystem multifunctionality increases with increasing chemodiversity, we first
102 exposed four levels of biodiversity (1, 2, 3, and 6 bacterial species) under controlled laboratory

conditions to three environmental sources of dissolved organic matter (DOM). The three DOM sources naturally varied in chemodiversity (1090, 1745, and 2564 formulae), which closely correlated with indices summarising potential biogeochemical functions, such as defined from the bioavailability, reactivity, and molecular mass of DOM[29] (Fig. S1). The molecular composition of the three DOM sources was also largely nested (Fig. S2), whereby increasing chemodiversity primarily consisted of the addition of aromatic-like formulae (Table S1). Our fully factorial design allowed us to compare the importance of chemodiversity versus biodiversity for predicting ecosystem multifunctionality from the main effects of chemodiversity and biodiversity and their interaction. Ecosystem multifunctionality was defined using changes in cell abundances, dissolved organic carbon (DOC) concentrations, and total dissolved nitrogen (TDN) concentrations over a 7-day period. We chose these functions because they are important measures of food web productivity and nutrient cycling[30]. Because high microbial activity can increase both DOM and TDN concentration by releasing exudates or transforming particulate organic matter into dissolved compounds, as well as reduce DOM and TDN concentration by assimilating and respiring organic compounds[31], we considered the absolute change in DOC and TDN concentration as an indicator of microbial activity. We found that these three functions were largely orthogonal to each other in our closed system, thereby representing complementary axes of ecosystem multifunctionality (Fig. S3). We then summed standardized values (*z*-scores) of the change in each of the three functions between day 0 and day 7 into a single index of ecosystem multifunctionality as is common in ecology[1,3]. Larger values of this index reflect higher ecosystem multifunctionality. We tested the main and interactive effects of chemodiversity and biodiversity as factors on multifunctionality using a linear model without any other variables, i.e. only accounting for the experimental design (see Methods).

128 We found strong experimental support for our hypothesis that chemodiversity enhances
129 ecosystem multifunctionality. Increasing chemodiversity led to higher values of both the
130 multifunctionality index (Fig. 1b) and each individual ecosystem function, with one exception
131 (Fig. S4). Over the range of observed chemodiversity (1090 to 2563 formulae),
132 multifunctionality was estimated to increase from a mean (95% confidence interval [CI]) of -
133 0.90 (-1.49 – -0.32) to a mean of 1.44 (0.10 – 2.77) (Fig. 1a). This mean increase corresponded
134 to an absolute change of 260% (107 – 966%). In support of these findings, DOM was
135 increasingly transformed with increasing chemodiversity. We specifically found ten-times as
136 many compounds were consumed at the highest compared to the lowest level of
137 chemodiversity, whereas the percentage of compounds that were produced in the incubations,
138 such as might arise from cell lysis, was relatively stable across chemodiversity treatments (Fig.
139 S5). Importantly, we found no evidence that chemodiversity was an outcome of greater
140 multifunctionality. The change in chemodiversity, that is, the difference between
141 chemodiversity measured at the end of the incubation in control incubations with no microbes
142 and chemodiversity measured at the end of the incubation in the presence of microbes, was not
143 correlated with multifunctionality (ρ=-0.22, p=0.203). Chemodiversity alone explained 24% of
144 the variation in multifunctionality, which is greater than the effect of biodiversity in many other
145 studies of ecosystem multifunctionality[3,8–10].
146
147 Neither ecosystem multifunctionality (Fig. 1b) nor any individual function increased
148 consistently with biodiversity in our experiment (Fig. S4). Although the two species
149 community had higher multifunctionality, its taxa were included in the six species community
150 (Fig. 1b). Thus, if increasing biodiversity increased ecosystem function, we would expect the
151 six species community to have an equal or higher multifunctionality than the two species
152 community. However, we observed lower multifunctionality with six species than with two

153 species (Fig. 1b). Adding biodiversity can sometimes have no added benefit for ecosystem
154 function, potentially because of competition and redundancy in the functions undertaken by
155 different species[32,33]. In our experiment, we found that the two species community grew faster
156 and caused larger absolute change in DOC and TDN than when its members were grown
157 alongside other taxa in the six species community (Fig. S4), suggesting competitive inhibition
158 in the latter. Most microbial taxa negatively interact when grown together[34], which reduces
159 their growth relative to expectations from when taxa are grown by themselves[35].
160
161 We also observed an interaction between chemodiversity and biodiversity that further
162 emphasizes the importance of chemodiversity for ecosystem multifunctionality. Higher
163 biodiversity increased ecosystem multifunctionality only at the highest chemodiversity level
164 (Fig. S6). These results suggest that without a sufficient diversity of compounds for microbes
165 to use, increasing biodiversity alone cannot sustain a higher ecosystem multifunctionality. By
166 contrast, in environments with low biodiversity and low chemodiversity, increasing
167 chemodiversity alone can increase ecosystem multifunctionality by enabling the species
168 already present to perform more functions. Because ecosystem multifunctionality-biodiversity
169 relationships[3,8–10,32,36–38] are mostly studied in terrestrial environments, where chemodiversity
170 is far higher than in aquatic environments[39], the importance of chemodiversity may have been
171 overlooked.
172
173 A potential criticism of our experiment is that because we observed an increase in absolute
174 carbon and nitrogen concentration, changes in ecosystem multifunctionality may have been
175 due to cell lysis rather than microbes transforming chemodiversity. However, both carbon and
176 nitrogen could have come from particulate organic matter as well as nitrogen fixation by some
177 of our study species[40,41]. In support of the former explanation, between 1.6 to 10.6% of the total

178 organic carbon in each treatment was in particulate rather than dissolved form. We also found
179 evidence that nitrogen was incorporated into DOM from other pools than cell lysis.
180 Specifically, we found that a higher proportion of N-containing compounds were produced at
181 higher chemodiversity, and this result could not have been due simply to cell lysis because
182 there was a higher overall consumption of compounds (Fig. S5). Taken together, our results
183 suggest that increasing chemodiversity increases ecosystem multifunctionality by enabling
184 microbes to use and transform substrates in more ways than is achieved by increasing
185 biodiversity alone.
186
187 To test the generality of our lab experiment, we sampled surface waters from 101 natural lakes
188 spanning continental Europe (~30° latitude). As in the lab experiment, the proportion of
189 aromatic compounds increased from 3.1% to 28.4% over the range of chemodiversity observed
190 in the field from 396 to 3387 molecular formulae (linear model, $t=6.54$, $df=100$, $p<0.001$). The
191 chemical characteristics of DOM, including its molecular weight, further suggested that
192 aromatic-like compounds were increasing with chemodiversity like in the lab experiment (Fig.
193 S7). We also measured different functions in the field than in our experiment to test if
194 chemodiversity promoted ecosystem multifunctionality irrespective of the specific functions
195 being measured, that is, if the association between chemodiversity and ecosystem
196 multifunctionality was generalizable. The functions measured in the field were related to
197 carbon cycling (CH_4 and CO_2 emissions) and ecosystem metabolism (total respiration,
198 maximum net primary production, and bacterial protein production). We reversed the sign of
199 the measured values for CH_4 emissions, CO_2 emissions, and total respiration so that a higher
200 multifunctionality value corresponded to a greater benefit to society[37] (see Methods). Thus,
201 larger values of these three functions indicated lakes were both emitting and mineralizing less
202 carbon. We then summed *z*-scores of the five functions into a single index as in the lab

203 experiment. We also compared the importance of chemodiversity with the importance of
204 microbial (bacteria, fungi, archaea) and eukaryotic biodiversity estimated from shotgun
205 metagenomic sequencing of environmental DNA in 82 of the study lakes. Our biodiversity
206 sampling detected more taxa than in other large-scale studies[42,43] with between 97 to 1589
207 species (median = 610) per lake.
208
209 We found a positive association between ecosystem multifunctionality and chemodiversity but
210 not with biodiversity across 82 of the 101 field sites with data for all these three variables. Over
211 the range of chemodiversity (396 to 3387 molecular formulae, median = 2445),
212 multifunctionality was estimated to increase from a mean of -1.27 (95% CI: -2.16 – -0.37) to
213 0.30 (-0.33 – 0.92) (linear model: t=2.24, p=0.030, df=81; Fig. 1c). This mean increase
214 corresponded to a change of 124% (43 – 144%). Neither biodiversity nor its interaction with
215 chemodiversity influenced multifunctionality (t=1.26, p=0.21 and t=-0.02, p=0.99,
216 respectively, df=81 for both). Much like in our experiment, the effect of biodiversity on
217 multifunctionality was limited by chemodiversity. We found that the effect of biodiversity on
218 multifunctionality only became statistically significant at the highest levels of chemodiversity
219 (Fig. S8). Three of the five measured functions, CH_4 and CO_2 emissions and bacterial protein
220 production (BPP), also varied individually with chemodiversity whereas only CO_2 emissions
221 significantly decreased with biodiversity (Fig. S9). At a molecular-level, these functions were
222 associated with different formulae, potentially reflecting differences in how microbial
223 functions consumed organic substrates. For example, BPP was negatively correlated with
224 generally low H:C formulae (Fig. S10), indicative of consumption[4], because these formulae
225 were relatively more abundant and thus easier for bacteria to find to build cellular structures
226 than rarer high H:C compounds that may release more energy[4,44]. CH_4 was positively
227 associated with the proportion of formulae per lake that contained either N or S (ρ=0.24,

228 p=0.006), which are precursors for oxic methanogenesis[45,46], and CO_2 negatively correlated
229 with high H:C formulae (Fig. S10), which are preferably respired by bacteria because they
230 provide the highest energy return[4,44]. Chemodiversity alone explained 5% of the variation in
231 multifunctionality, which is lower than in our experiment but we considered more functions
232 than in our experiment and our experiment controlled for environmental noise, e.g. climate.
233 This result is also comparable to, and in some occasions up to five-times higher than, the effect
234 of biodiversity reported in influential field studies explaining variation in ecosystem
235 multifunctionality[8,36]. The choice of functions affects the strength of the relationship between
236 biodiversity and multifunctionality, so may explain some of these differences among studies.
237
238 To test further the importance of chemodiversity relative to published estimates of biodiversity,
239 we computed multifunctionality based on the exceedance of data-derived thresholds[37]. This
240 method was specifically designed to facilitate inter-study comparisons[37]. We found that the
241 reduced explanatory power of biodiversity relative to chemodiversity persisted when we
242 estimated ecosystem multifunctionality with this threshold approach (Supplementary Note 1).
243 Our estimated effect of chemodiversity on multifunctionality with this threshold approach was
244 larger than the effect of biodiversity on ecosystem function that was previously estimated from
245 47 published studies[9], including freshwaters (Supplementary Note 1). Unlike for biodiversity-
246 ecosystem multifunctionality, where positive associations have been reported in studies
247 measuring plant productivity[3,8], and rarely in studies measuring macronutrient cycling[33], the
248 chemodiversity-ecosystem multifunctionality relationship may be less sensitive to the specific
249 functions measured. We found consistent positive associations between chemodiversity and
250 ecosystem multifunctionality across our experimental and field datasets with fundamentally
251 different functions associated with primary productivity and macronutrient cycling. Together,
252 our results from both controlled and natural conditions suggest that the large role of

253 chemodiversity in supporting ecosystem multifunctionality has, to date, been overlooked and
254 must be considered in efforts aimed at delivering benefits from nature to society.
255
256 *Recalcitrance underlies positive chemodiversity-multifunctionality associations*
257 Based on theoretical models that predict that the reactivity of DOM varies with its molecular
258 composition and local environmental conditions[13,27], we hypothesised that the effect of
259 chemodiversity on ecosystem multifunctionality depends on the source of chemodiversity
260 itself. By sampling lakes at similar periods of annual heat accumulation, and with a similar
261 light environment and DOC concentration (see Methods), we found that spatial variation in
262 chemodiversity at a continental scale primarily reflected differences in climate and surrounding
263 land use[16,47] (Fig. S11). Specifically, chemodiversity strongly increased from southern to
264 northern Europe (ρ=0.40, $p<0.001$), suggesting that the environment in which chemodiversity
265 occurred and/or its source could explain why chemodiversity varied throughout Europe. Here
266 we extend these two potential explanations (see Supplementary Note 2 for further discussion)
267 to contend that they differentially influence how chemodiversity is used by microbes and thus
268 ecosystem multifunctionality. If chemodiversity was high because of environmental conditions
269 ("environment" hypothesis in Fig. 2a) that slow microbial transformation of DOM, such as
270 lower mean annual temperature (MAT) or lower macronutrient concentrations, we expected
271 that ecosystem functions like microbial respiration and bacterial protein production would also
272 be reduced[48], resulting in a negative correlation between chemodiversity and
273 multifunctionality. By contrast, if chemodiversity primarily arose because compounds with
274 greater chemical complexity were imported into lakes, we expected these compounds to
275 support a greater number of biochemical transformations and increase ecosystem
276 multifunctionality ("source" hypothesis in Fig. 2a). Complex compounds, such as those with
277 high aromaticity, are abundant in lakes and their degradation by microbes involves many

278 biochemical pathways that increase metabolic activity and biomass formation[4,49,50], and
279 ultimately multifunctionality. By contrast, simpler compounds are expected to decrease
280 multifunctionality because they are often used to produce energy, which emits CO_2 as a by-
281 product.
282
283 We tested how both the complexity of DOM and the environment in which DOM was degraded
284 affected microbial transformation of DOM and subsequently the effect of chemodiversity on
285 ecosystem multifunctionality. We estimated DOM complexity in each lake based on an
286 aromaticity index[51] (see Methods). We expected this index to reflect terrestrial vegetation[16],
287 which is a main source of lake DOM[40]. Slower growing vegetation tends to invest more in
288 protecting biomass, such as with secondary metabolites[52], because it faces a relatively greater
289 cost to produce biomass. Compounds involved in plant protection tend to be more complex and
290 have chemical properties, including a higher aromaticity, which require more biochemical
291 reactions to be transformed[14,27]. We expected complexity to influence multifunctionality by
292 determining the extent to microbes can transform DOM and we expected complexity itself to
293 be determined by terrestrial net primary productivity (NPP), i.e., vegetation growth rates, in
294 surrounding catchments. We also considered nine environmental variables (MAT, annual
295 precipitation, water temperature, pH, water colour, lake area, biodiversity, trophic state, and
296 nutrient concentrations) that could influence chemodiversity by changing the diversity of
297 terrestrial and aquatic organic matter inputs, controlling the physical energy required to break
298 chemical bonds, and regulating photo-oxidation[53,54,55]. NPP and the environmental variables
299 were estimated from high-resolution remote sensing data and in-situ measurements (see
300 Methods). We then averaged the number of times molecular formulae could be transformed in
301 each lake using a list of 1255 putative biochemical transformations[19] (see Methods). If the
302 positive association we observed between chemodiversity and ecosystem multifunctionality

303 arose from the input of complex compounds rather than the limitation of microbial
304 transformations of DOM, we expected NPP would be more strongly associated with
305 chemodiversity than the nine other environmental variables.
306

307 Our results indicate that the source of DOM explains how and where a positive chemodiversity-
308 ecosystem multifunctionality relationship can arise. Specifically, we found that the production
309 of complex compounds by low terrestrial NPP (ρ=-0.70, $p<0.001$, Fig. S12a) correlated with
310 both greater chemodiversity (ρ=0.56, $p<0.001$; Fig. S12b) and a greater number of average
311 transformations that formulae could undergo in each lake (ρ=0.45, $p<0.001$, Fig. S12c).
312 Consequently, chemodiversity correlated positively with the number of transformations per
313 formulae (ρ=0.40, $p<0.001$, Fig. 2b). Correlations between chemodiversity and each of
314 complexity, NPP, and the number of biochemical transformations are predicted by the source
315 hypothesis. By contrast, all the environmental variables had weaker associations with
316 chemodiversity, though some were statistically significant (Table S2). Together, our results
317 indicate that the production of complex compounds by low NPP increases both chemodiversity
318 and the number of biochemical transformations that each compound can support.

319

320 We synthesised the observed correlations among chemical complexity, biochemical
321 transformations, and chemodiversity to propose a new general understanding of ecosystem
322 multifunctionality. This new understanding was that the positive effect of chemodiversity on
323 multifunctionality arises because of the production of complex compounds by plants and
324 environmental conditions that favour those compounds within DOM pools. We found further
325 support for this framework with a path analysis explicitly testing presumed causal pathways.
326 The path analysis modified our original hypotheses to incorporate additional paths identified

327 by tests of model structure (Fig. S13). With this analysis, we found that effects related to the
328 production of organic compounds (i.e. NPP) influenced chemodiversity by changing DOM
329 complexity and subsequently the number of biochemical transformations it can receive (Fig.
330 2a). Although environmental variables (i.e., water colour and MAT) also influenced DOM
331 complexity, the direction of these effects suggested that they were associated with reduced in-
332 lake primary productivity and an increase in the relative importance of terrestrial DOM (Fig.
333 S13). This interpretation was further supported by the direct negative effect of chlorophyll *a*
334 concentrations on chemodiversity (Fig. S13).

335

336 To test further if ecosystem multifunctionality was enhanced by chemodiversity because of the
337 properties of complex compounds associated with low terrestrial NPP, we analysed the
338 composition and putative source of all 8456 molecular formulae identified in the field survey.
339 We first identified formulae whose relative abundances across our study lakes were correlated
340 with terrestrial NPP of the surrounding catchment. Positive and negative correlations suggest
341 formulae likely originated from high and low terrestrial NPP, respectively. We compared these
342 correlations with the average number of biochemical transformations that each formula was
343 predicted to undergo in each lake[19]. This analysis allowed us to infer the origins of compounds
344 that can be highly transformed by biotic and abiotic reactions, which contribute to higher
345 ecosystem multifunctionality. Compounds that were associated with low NPP appeared more
346 transformed ($t=-8.19$, $df=6639$, $p<0.001$), and positively correlated with chemodiversity
347 ($t=8.50$, $df=6639$, $p<0.001$; warmer colours in Fig. 2c). These compounds were more complex,
348 on average, with a mean (± standard error) aromaticity index of 0.52 (±0.17) compared to 0.32
349 (±0.24) in the entire dataset (Wilcoxon rank sum test, $W=5.4 \times 10^6$, $p<0.001$, Fig. S14). These
350 results further supported the hypothesis that complex compounds can be degraded in more
351 ways and support higher multifunctionality. We then selected molecular formulae that most

352 increased in abundance with decreasing NPP, that is, were in the top 10% of negative
353 associations based on Spearman rank correlations. We found that these formulae could support,
354 on average, 1.4 times more known putative biochemical transformations than formulae that
355 were weakly associated with NPP, i.e. lowest 10% of negative associations ($t=5.80$, $df=860$,
356 $p<0.001$), and 1.8 times more transformations than formulae strongly positively associated with
357 NPP, i.e. top 10% of positive correlations ($t=8.00$, $df=1090$, $p<0.001$). Together, these results
358 demonstrate that the effect of chemodiversity on multifunctionality emerges from the
359 properties of individual formulae.
360
361 We expect chemodiversity will be less important for ecosystem multifunctionality in
362 environments where microbes cannot access and transform large quantities of organic matter.
363 For example, DOM concentrations are typically >5 mg C L^{-1} in rivers, most lakes, wetlands, or
364 mangroves[19,47,48,56], but <1 mg C L^{-1} in the deep ocean[29] or in large (surface area >500 km^2)
365 deep lakes with centennial residence times[57]. At these concentrations, individual compounds
366 are more likely to be too dilute to interact with microbes[52], irrespective of the diversity of
367 compounds in the environment, and may result in a shallower slope between chemodiversity
368 and ecosystem function. In support of this expectation, we found that compounds transformed
369 in lakes reacted less frequently when found in the deep ocean. We identified 2745 formulae
370 shared between our study lakes and a representative deep ocean sample[18], which is likely an
371 overestimate because we cannot differentiate structural isomers. We then calculated the
372 number of times that each formula was transformed in each sample using the established list
373 of putative biochemical transformations[19]. We found that these formulae underwent, on
374 average, 0.08 (range: 0 to 1.32, standard deviation = 0.17) putative transformations in lakes but
375 no transformations were detected in the deep ocean. Given that many of the detected
376 transformations are ubiquitous, such as methylation, a likely explanation for why we did not

377 find them in the ocean sample is because their substrates occur at concentrations lower than
378 our femtomole detection limit[58]. Such concentrations would be several order of magnitude
379 beneath concentrations of the most limiting organic compounds[59]. This observation further
380 suggests that chemodiversity will have a greater influence on ecosystem multifunctionality
381 where individual compounds occur at high enough concentrations to support many biochemical
382 transformations.
383
384 *Chemodiversity and ecosystem multifunctionality under global change scenarios*
385 NPP is expected to increase because of rising atmospheric CO_2 concentrations[60], and our results
386 suggest that this increase could reduce the chemodiversity of lakes in ways that are relevant to
387 global carbon budgets and climate modelling efforts. To predict how the resulting changes in
388 chemodiversity will influence ecosystem multifunctionality, we first modelled the value of
389 each ecosystem function that varied with chemodiversity in our field survey (Fig. S9). By
390 adding chemodiversity to statistical models of CH_4 and CO_2 emissions and BPP that contained
391 predictors identified as important in previous large-scale studies[23,24], we increased explained
392 variation of each function by up to 50% as compared with past studies (Table S4). We achieved
393 the large increase in explained variation despite being unable to measure some prominent
394 drivers of ecosystem metabolism, such as phosphorus concentrations[61]. Importantly, based on
395 the properties of linear models, additional predictors are unlikely to change the effect of
396 chemodiversity on individual functions and will only further reduce unexplained variation. We
397 then projected our models for each ecosystem function to all European lakes >10 ha (see
398 Methods). For chemodiversity, we generated a European-scale raster with the model used in
399 the path analysis (Fig. S13). This analysis further validated our study because we found that
400 the extrapolated estimates were within bounds of global-scale studies[23,24]; no European-only
401 studies exist for comparison. Overall, we estimated that European lakes emitted, on average,

402 22.7 Tg C yr^{-1} as CH_4 (95% CI 11.2 – 32.0; model R^2 = 0.46) and 27.8 Tg C yr^{-1} as CO_2 (95%
403 CI: 9.9 – 37.4; R^2 = 0.22) in 2019. Despite these emissions potentially being overestimated
404 because they only considered summertime fluxes[62], they corresponded to 7% and 19% of
405 estimated global lake CH_4 and CO_2 emissions[23,24], respectively, and Europe contains ~20% of
406 the world's total number of lakes[63]. We also estimated that European lakes fixed an average of
407 1.8 Tg C yr^{-1} (95% CI: 1.0 – 3.3; R^2 = 0.45) in 2019 into microbial biomass, that is, from BPP.
408
409 To forecast the consequences of changes in NPP for European lakes, we input NPP estimated
410 in the year 2100 by Earth system models[64] into the linear model that predicted chemodiversity
411 (Table S3) and then used the predictions to forecast each ecosystem function (Table S4). We
412 focused solely on changes in NPP, holding other variables constant, because NPP was the best
413 individual predictor of chemodiversity (Table S2) and had a clear mechanistic link to
414 chemodiversity (Fig. S13). We used two shared socioeconomic pathways (SSPs) representing
415 likely (SSP2) and worst case (SSP5) scenarios for future changes in atmospheric CO_2
416 concentrations. We found that future increases in NPP alone will reduce chemodiversity in
417 99.1% of lakes, with the median chemodiversity decreasing by 13% and 24% under the SSP2
418 and SSP5 scenarios, respectively (Fig. 3a). When summed across lakes, we estimated that the
419 change in chemodiversity will cause lakes to emit an additional mean 28.4 (95% CI: 3.3 –
420 166.1) and 32.2 (95% CI: 14.0 – 231.2) Tg C yr^{-1} as CH_4 under SSP2 and SSP5 scenarios,
421 respectively (Figs 3b, S15). Similarly, the changes in chemodiversity will cause lakes to emit
422 an additional 3.7 (95% CI: 1.3 – 5.1) and 4.0 Tg C yr^{-1} as CO_2 (95% CI: 1.3 – 5.2 Tg C yr^{-1})
423 under SSP2 and SSP5 scenarios, respectively (Figs 3c, S15). Microbes will, however, fix 4.3
424 (95% CI: 1.7 – 10.2) and 4.4 Tg C yr^{-1} (95% CI: 1.7 – 12.6 Tg C yr^{-1}) more carbon into biomass
425 under SSP2 and SSP5 scenarios, respectively (Figs 3d, S15). Together, our estimates suggest
426 that European lakes may emit, on average, 1.6-times more carbon by 2100, mostly as CH_4,

427 because of NPP-associated declines in chemodiversity. Although our estimates are higher than
428 previous estimates of future CH_4 emissions by lakes in absolute terms by 50%, they are
429 comparable in terms of relative change, i.e. 73-84% in our study versus 58-88%[62]. The
430 difference in absolute terms is likely due to the lack of winter measurements in our dataset as
431 lower microbial activity in colder weather typically reduces greenhouse gas emissions[62]. The
432 similarity in relative terms instead emphasises that chemodiversity is a necessary predictor of
433 CH_4 emissions.
434
435 We forecasted that lakes in southern Europe, and in the south of the Scandinavian peninsula,
436 will experience the largest changes in ecosystem function (Fig. 3). These lakes are located near
437 large population centres and, therefore, provide opportunities to improve management of their
438 surrounding catchments in ways that can enhance the benefits that European lakes provide to
439 society. For example, our results suggest that chemodiversity and subsequently ecosystem
440 multifunctionality can be enhanced by promoting slow growing vegetation, such as many
441 Fabaceae[65]. Slow growing vegetation produce large numbers of complex polymeric
442 compounds, which undergo many biochemical transformations when leached into water from
443 litterfall and soils[17]. They also have higher wood density[66], that can increase long-term carbon
444 storage on land[67], resulting in a win-win for land and water.
445
446 *Conclusions*
447 Here, we provided experimental and large-scale observational evidence that increasing
448 chemodiversity delivers beneficial ecosystem functions, especially a stronger carbon sink
449 capacity corresponding with greater bacterial biomass and lower greenhouse gas emissions.
450 We then demonstrated that the ways in which ecosystems deliver these functions depends on
451 the molecular properties of DOM, namely the capacity of individual compounds to undergo

452 biochemical transformation. The molecular properties of compounds ultimately depend on
453 sources of DOM and external environmental conditions, such as their concentrations, as found
454 in many inland waters, wetlands, and mangroves[19,47,48,56]. Here we add to this understanding
455 by showing the importance of these properties for ecosystem multifunctionality and their
456 potential to explain ecosystem-specific differences in chemodiversity–multifunctionality
457 relationships. Our findings also suggest that chemodiversity and its effect on multifunctionality
458 can be altered across different aquatic ecosystems by manipulating sources of DOM within
459 surrounding catchments. Finally, our study suggests that accounting for the strong association
460 between NPP and chemodiversity can improve integration between terrestrial and aquatic
461 ecosystem models[5]. For example, because chemodiversity is strongly predicted by NPP, which
462 itself is predictable at high temporal and spatial scale from bioclimatic data, we can readily
463 incorporate the effect of chemodiversity on greenhouse gas emissions into existing carbon
464 budgets. More generally, our study demonstrates that sustaining the benefits that lakes provide
465 to people will require management that extends beyond shorelines and considers the
466 chemodiversity exported from land.

467

468 **Acknowledgments**

469 We are grateful for help with field work from Samuel Cottingham, Carolyn Ewins, Sophie
470 Guillaume, Eleanor Sheridan, and Samuel Woodman. We thank Yi Zhang and Caroline Kemp
471 for laboratory assistance, Del Hawtin, and Simon Hoer for logistical support, and Katrin
472 Klaproth for assistance with FT-ICR MS measurements. We thank Jim McGeer for providing
473 the DOM sources and Thomas Scheuerl for providing the bacterial strains used in the
474 experiment. This work was funded by H2020 ERC Grant sEEIngDOM #804673 to AJT and
475 the University of Cambridge Harding Distinguished Postgraduate Scholars Programme.

476

477 **Author contributions**

478 AJT and JAF designed the study. JAF performed the field and associated lab work with input

479 from AJT and TD. SRS and JAF performed the laboratory experiment. JAF analysed the data

480 with input from AJT. JAF and AJT wrote the manuscript with input from TD and SRS.

481

482 **Data availability**

483 All data are available within the manuscript or upon reasonable request.

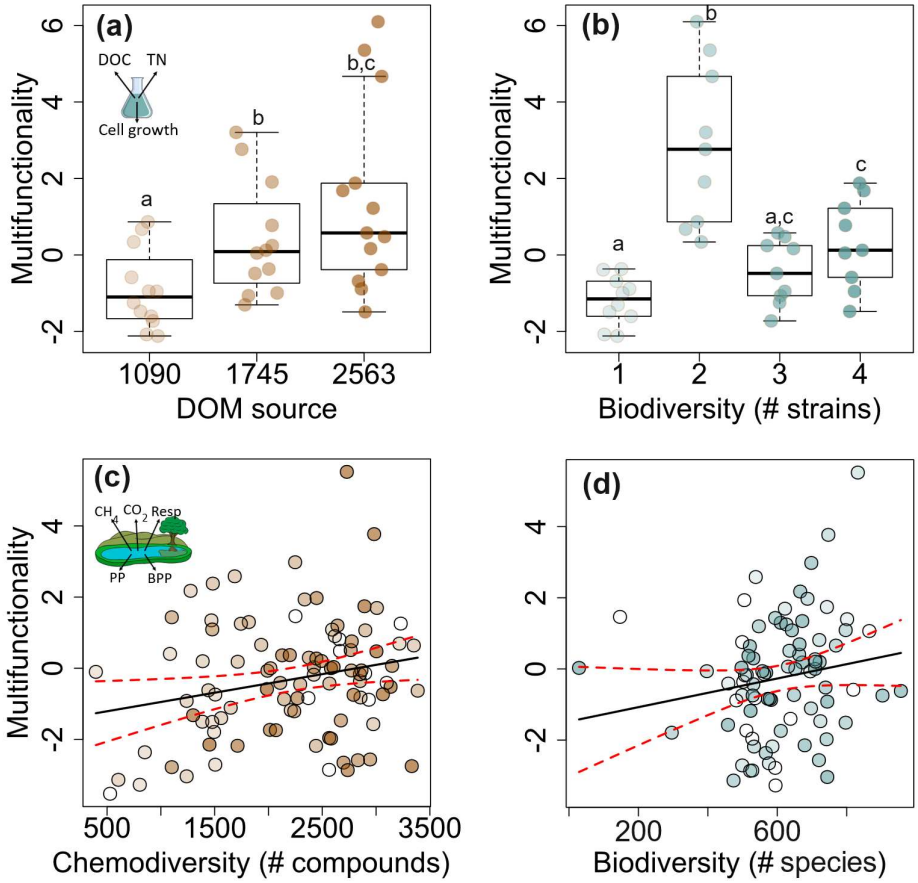

Figure 1: Chemodiversity increases ecosystem function. Multifunctionality increased with **(a)** higher chemodiversity but not with **(b)** higher biodiversity under controlled conditions. Multifunctionality was measured as the sum of *z*-scores of changes in each of cell abundances, and transformation of dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) over 7 days in 36 microcosms represented by individual points. Boxes encompass the inter-quartile range with median denoted by solid horizontal line and whiskers equal to 1.5-times the interquartile range. Different letters indicate statistically significant ($p<0.05$) differences between factor levels with Tukey's pairwise comparison tests. Chemodiversity of each experimental source was estimated from the exponential of the Shannon diversity index. In a separate field study, multifunctionality increased with **(c)** higher chemodiversity but not with **(d)** higher biodiversity across 82 European lakes represented by individual points. Multifunctionality was measured in the field study as the sum of *z*-scores of CH_4 and CO_2 emissions, total respiration (Resp), maximal net primary production (PP), and bacteria protein production (BPP). Lines in **(c)** and **(d)** are estimated mean associations based on a linear model ± 95% confidence intervals. For all, darker colours indicate higher chemodiversity (brown points) and biodiversity (blue points).

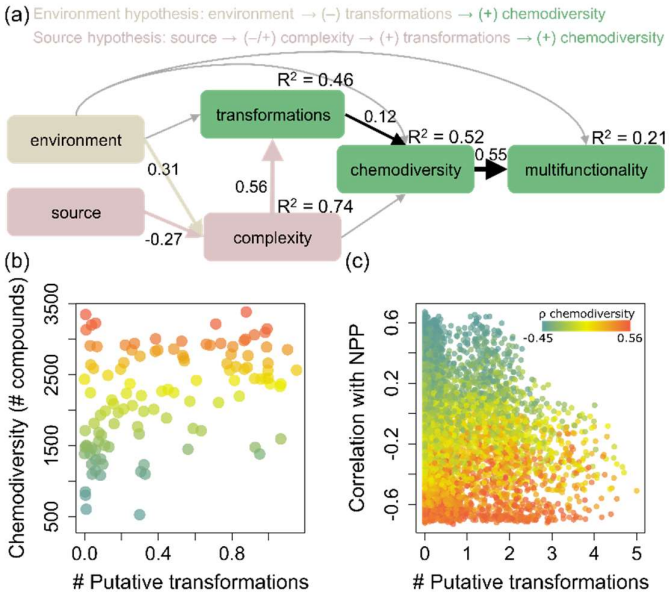

Figure 2: The origin of chemodiversity influences its effect on ecosystem multifunctionality. (a) We hypothesized that the positive effect of chemodiversity on ecosystem multifunctionality was caused by environmental conditions reducing the transformation of DOM ("Environment" hypothesis) and sources of DOM producing compounds that could be transformed in a greater number of ways ("Source" hypothesis). + and – symbols show hypothesised cause-and-effect of one variable on another, with either positive or negative links between NPP and complexity possible in the Source hypothesis. We tested these hypotheses using path analysis fitted to 101 lake-level observations. To simplify the path analysis, we created a composite variable "environment" that encompassed 9 measured parameters (see Fig. S13 for effect of each parameter), whereas "source" was equal to NPP in the surrounding catchment of each lake. Complexity and transformations corresponded with the intensity-weighted modified aromaticity index and mean number of putative transformations per molecular formula per lake, respectively. Coloured arrows show statistically significant paths with standardized effects. **(b)** The mean number of transformations per formula per lake was closely correlated with chemodiversity (ρ=0.40, $p<0.001$). Each dot represents a lake (N=101) coloured for chemodiversity. **(c)** We also correlated the number of transformations that each of 8456 molecular formulae underwent across all lakes with the NPP of the lake catchment. Formula (points) associated with low NPP (i.e., negative value along the *y*-axis) generally underwent more transformations (i.e. larger value along the *x*-axis; ρ=-0.10, $p<0.001$). Numbers of transformations were weighted by the relative abundance of each formula across lakes. Points are coloured according to their correlation with chemodiversity with low (blue) and high (red).

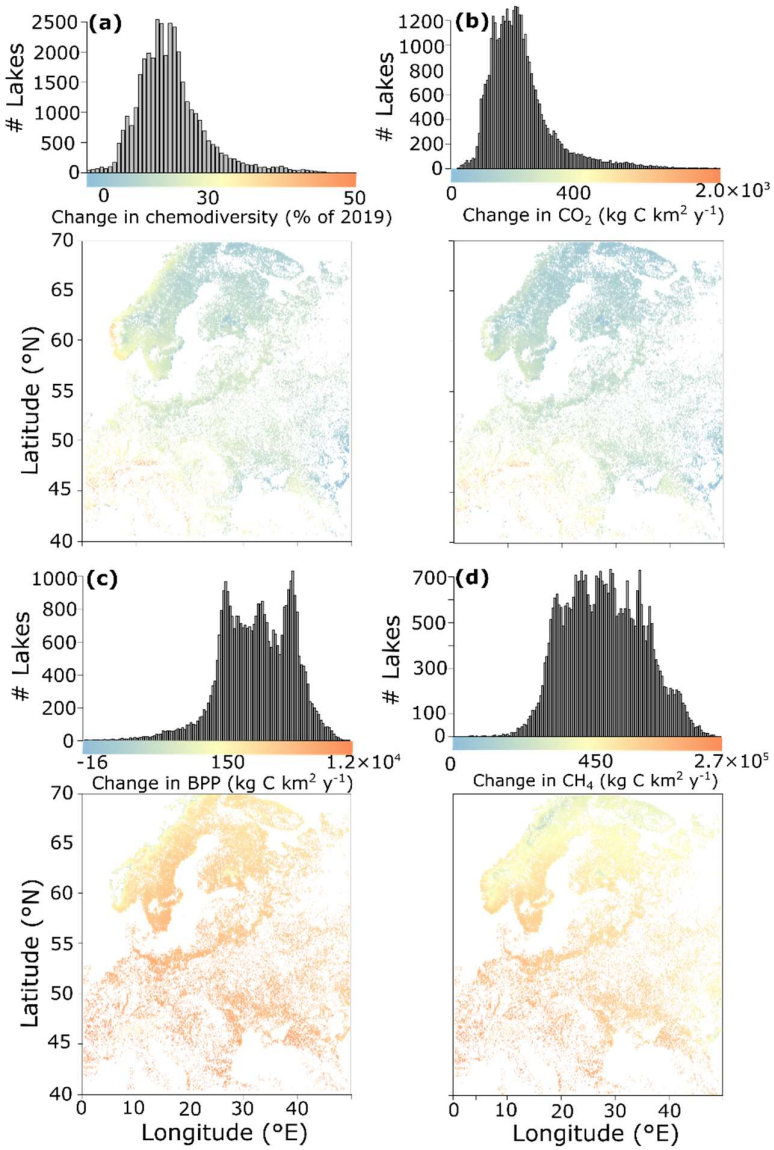

Figure 3: Lakes will emit more carbon by 2100 due to continental-wide declines in chemodiversity. We estimated **(a)** chemodiversity in 2100 while holding all other variables at 2019 levels for all European lakes >10 ha in surface area (R^2 for chemodiversity model = 0.36). Values are the change in chemodiversity compared to 2019. Using future chemodiversity, we then estimated mean changes in **(b)** CO_2 emissions **(c)** bacterial biomass production, and **(d)** CH_4 emissions from pelagic surface waters between 2019 and 2100 using linear models (R^2 = 0.22, 0.45, and 0.46, respectively; Table S4). For all, points indicate N=175,500 individual lakes. Histograms indicate the number of lakes found with corresponding changes. 95% confidence intervals for each function are given in Fig. S15.

534 **Materials and methods**

535 *Laboratory experiment*

536 To test if multifunctionality varies with chemodiversity, we incubated bacterial communities
537 of four biodiversity levels with three different levels of chemodiversity. The four biodiversity
538 levels were 1) *Pseudomonas fluorescens* SBW25 2) *Chryseobacterium* sp. and
539 *Sphingobacterium* sp. isolated from natural communities that degrade DOM[68]; 3) *Rahnella* sp.,
540 *Raoultella* sp., and *Pantoea* sp. from the same natural communities[68]; 4) all six microbes
541 together. Before starting the experiment, we grew each microbial isolate individually from a
542 frozen archive for five days in R2A media. A control of filtered (0.2 µm, Sterivex-GP,
543 MilliporeSigma, USA) sterile water was included alongside each treatment and indicated no
544 contamination as measured with flow cytometry (see below). For chemodiversity, the three
545 treatment levels were obtained from reverse-osmosis-concentrated water[69] of a temperate
546 wetland (Luther Marsh, Ontario, Canada; 43°57'43”N, 80°23'50”W), boreal lake (Clearwater
547 Lake, Ontario, Canada; 46°22'11”N, 81°03'04”W), and Arctic river (Cameron River,
548 Northwest Territories, Canada; 62°29'33”N, 113°32' 48”W). Using MilliQ water, we diluted
549 each reverse-osmosis concentrate to a final dissolved organic carbon (DOC) concentration of
550 10 mg C L^{-1}, as measured with a TOC-TNM-L analyser (Shimadzu, Japan), and filled fifteen
551 250 mL pre-combusted (4 hr at 550°C) Erlenmeyer flasks with 100 mL of each water source
552 (n = 3 replicates for each of 4 biodiversity levels with 3 controls). Before inoculating the DOM
553 sources with each microbial community, we estimated strain-specific cell number with a
554 CLARIOstar fluorometer (BMG Labtech, UK) after 30 min of staining with SYBR Green
555 (10,000× stock solution, Invitrogen, USA) diluted to a 10× final concentration. Based on these
556 data, we diluted an unstained aliquot of each strain to an expected 10,000 fluorescence units
557 and mixed them to form the four communities. This normalisation created a substitutive design
558 that ensured each community had a similar abundance of DNA and thus potential activity.

559 Then, we added 1 mL of each of the four microbial communities to each chemodiversity level
560 in triplicate and incubated the flasks at ca. 20°C for seven days in the dark without shaking.

561

562 *Measuring chemodiversity*

563 We estimated chemodiversity in the experiment from solid-phase extraction of DOM following
564 a standard protocol[70]. Briefly, 80 mL of each incubation flask was acidified to a pH of 2 by
565 adding HCl, passed through 1 g of a polar styrene-divinylbenzene polymer (Bond Elut PPL,
566 Agilent, USA), and eluted with 3.5 mL of ultrapure methanol (LC-MS LiChrosolv, Merck,
567 Germany). We diluted the extracts to a final concentration of 2.5 ppm DOC in a 1:1 (v:v)
568 methanol:water solution before injecting 100 µL into a 15 Tesla Solarix FT-ICR MS (Bruker
569 Daltonic, Germany) with direct infusion electrospray ionization in negative mode. We
570 collected 200 scans in duplicate for each sample and calibrated the spectra using the
571 DataAnalysis software (Bruker Daltonic, Germany). Masses ranging from 150 to 1000 m/z
572 were exported, and we assigned molecular formulas using the online platform ICBM-
573 OCEAN[71]. We used the N, S, P rule, deleted all singletons, and used isotopic confirmation to
574 accept/reject a formula. We limited formula attributions to $C_{1\text{-}100}H_{2:200}O_{0\text{-}70}N_{0\text{-}3}P_{0\text{-}2}S_{0\text{-}2}$ with a
575 tolerance of 0.2 ppm and retrieved a total of 3873 individual molecular formulas. We computed
576 chemodiversity as the exponential of the Shannon diversity index using all assigned molecular
577 formulas[72] and the summed normalised peak intensity per sample. This index represents the
578 diversity in a community when all formulae are equally common, so corrects for variation in
579 evenness that can bias raw counts[73] of formulae number. We also computed the ratio of
580 hydrogen to carbon atoms in each compound as an indicator of bioavailability, as described
581 elsewhere[44]. The three DOM sources were largely nested with one another with 93% of
582 formulae present in the Arctic river (low chemodiversity) also present in the boreal lake
583 (medium chemodiversity), 70% of boreal lake formulae present in the temperate wetland (high

584 chemodiversity), and 80% of formulae present in the Arctic river present in the temperate
585 wetland (Fig. S2). The large number of shared formulae among treatments confirmed that
586 sources with higher chemodiversity contained the vast majority of formulae present in sources
587 with lower chemodiversity, i.e. were nested. Additionally, the molecular composition of the
588 DOM used in the experiment differed as much from the lakes sampled in the field study as the
589 lakes differed among themselves (Fig. S16).

590

591 *Ecosystem multifunctionality in the laboratory experiment*

592 We estimated the multifunctionality of each replicate community in the experiment by
593 measuring carbon consumption, nitrogen consumption, and bacterial growth rate. At the start
594 and end of the incubation, we measured the DOC and TDN concentrations of each flask using
595 a Shimadzu TOC-TNM-L analyser. Carbon and nitrogen consumption were estimated as the
596 change in concentration between the start and end of the incubation. To measure growth rate,
597 we determined the difference in cell abundance between the start and end of the incubation
598 using flow cytometry. Briefly, we removed 800 µL of water from each flask at the start and
599 end of the experiment. Samples were incubated with 200 µL of 25% glutaraldehyde in the dark
600 at 4°C for 30 min before being preserved at -80°C. Within one week, we thawed the samples,
601 added SYBR Green at a 10× final concentration (10,000× stock solution, Invitrogen, USA),
602 and incubated the samples overnight at 4°C. Then, we injected 100 µL of each sample into a
603 flow cytometer (Attune NxT, Thermo Fisher Scientific, USA) at a rate of 250 µL per minute
604 and used the same instrument settings for all samples. We optimised the gating for each DOM
605 source based on stained/unstained samples using FlowJo v10.9 (BD Biosciences, USA) and
606 used the same gating for each community. Each function was standardized to a mean of zero
607 and standard deviation of one to obtain a *z*-score, which we summed across the three functions
608 to estimate multifunctionality[8].

Selection of field study sites

We selected lakes to sample across Europe ensuring there were no systematic biases in DOC concentrations and water chemistry across our latitudinal gradient. Using the EU's Water Information System for Europe (WISE) database version 4[74], we subsampled all entries labelled as "lakes" for which DOC was reported in mg C L^{-1}. We also selected entries with a total organic carbon (TOC) concentration equal or lower than 5 mg C L^{-1} to include more lakes with low DOC values, which were underrepresented in the WISE database despite being known to be the most common on Earth[75]. From the 2169 resulting entries, we removed 256 entries with a DOC/TOC concentration >30 mg C L^{-1}, which is outside the expected values in lakes[75] and likely represent measurement error or atypical lakes, and 1595 duplicated measurements. We grouped the remaining 318 lakes into five bins in 5 mg C L^{-1} intervals: <5 mg C L^{-1}, 5-10 mg C L^{-1}, 10-15 mg C L^{-1}, 15-20 mg C L^{-1}, >20 mg C L^{-1} and six 5° latitudinal intervals: <45°N, 45-50°N, 50-55°N, 55-60°N, 60-65°N, >65°N. The last two DOC/TOC bins (i.e. >15 mg C L^{-1}) were absent from the first two latitudinal bins. Therefore, we removed all 47 entries with DOC/TOC concentrations >15 mg C L^{-1}. In a global analysis[75] of lake DOC concentrations, 93% of lakes were predicted to have DOC concentrations <15 C L^{-1}, so this filtering minimises potential sampling bias while ensuring we still represent the vast majority of lake DOC concentrations. To standardise the sites further, we considered total nitrogen (TN) and total phosphorus (TP) concentrations of all entries. Six entries lacked TP values, so we discarded all but one, which was located between 50-55°N where entries were sparse. Only 10 entries had TP concentrations higher than 0.2 mg L^{-1}, and 21 entries had a TN value higher than 1 mg L^{-1}, so we removed these to avoid large differences in lake trophic levels.

633 After reducing the number of entries to lakes with similar water chemistry, entries were
634 selected based on their distance from one another. Therefore, we discarded 28 entries from
635 eastern Poland, one in Latvia, and two in Hungary as they were logistically challenging to
636 reach. We similarly removed 175 entries situated in the United Kingdom and western France.
637 We then inspected each remaining entry individually using Google Earth satellite imagery,
638 removing 86 sites that were dammed or visibly artificial. This process removed all entries in
639 the WISE database situated south of 40° N and north of 69°N. To extend the latitudinal
640 gradient, we manually added eight sites in southern Italy (<40°N) and seven in northern
641 Norway (>69°N) without knowledge about their DOC/TOC or trophic state. None of the
642 manually chosen lakes were dammed or visibly artificial. This selection process resulted in a
643 total of 109 lakes. During sampling, severe weather conditions, unexpected fences around
644 lakes, or equipment failure prevented us from sampling eight lakes, and one site in Poland was
645 substituted for a lake for which no DOC data were available. Ultimately, we sampled 101 lakes
646 spread over 31 latitudinal degrees. There was no statistically significant correlation between
647 latitude and DOC concentration at the time of sampling (Spearman correlation, ρ=0.05, p=0.57,
648 Fig. S17).
649
650 *Timing of field sampling*
651 The order in which lakes were sampled was chosen to minimise differences in annual heat
652 accumulation and water temperatures across the latitudinal gradient. Sampling dates were
653 originally chosen to minimise differences in growing degree days (GDD) among lakes. GDD
654 represents the difference between a given temperature threshold at which primary production
655 commences and mean daily temperature[76]. Here we used a threshold of 10°C as is classically
656 employed[76]. To compute GDD, we used the rnoaa package[77] in R v.4.1.1[78] to retrieve all
657 weather stations within 50 km of the 101 sampled lakes from the US National Oceanic and

Atmospheric Administration National Climatic Data Center database. We determined the median GDD for each day of the year based on temperature data collected from the last 10 years. For years with 366 days, the GDD on the 29th of February was added to the GDD from the 28th of February. Cumulative GDD (cGDD) – representing the sum of GDD from the 1 January until a given date – were obtained by sequential addition for each day of the year for each lake. The computed cGDD indicated that the growing season commenced (i.e. cGDD >10) as early as the 1 January at the southernmost site and ended (i.e. cGDD reached its highest values and no longer increased) as early as the 10 October at the northernmost site.

We then selected sampling dates to survey lakes during similar periods of annual heat accumulation. To find the optimal sampling day, we calculated the variation in cGDD along the latitudinal gradient as the ratio between the 3rd and the 1st quartiles of lake-level cGDD, assuming we started sampling the southernmost site and moved northwards to the closest site at a pace of one lake per day. By starting sampling on the 8 April 2019, we could minimise the variation in cGDD across the latitudinal gradient. Overall, the observed maximal difference between the 5° latitudinal bins was 194.4 GDD, which was much less than the 368 that was originally expected based on the historical data (Fig. S17). This value corresponded with sampling the warmest site 10 days later than the coldest site, during which temperatures remained above 20°C. We further measured underwater irradiance at 1 m depth in 60 lakes across the latitudinal gradient using a HOBO light pendant (UA-002-64, Onset, USA) to test for additional biases in irradiance. Pictures representing the latitudinal gradient are presented in Fig. S18.

Dissolved organic matter sampling

682 At each lake, we collected 10 L of surface water in an acid-washed plastic carboy. We sampled
683 from the point with maximum water column depth or the point equidistant to all shorelines
684 when no bathymetry information was available. We expected these locations would be the best
685 mixed in the epilimnion and most representative of the lake water quality. Directly after
686 sampling, 1 L of lake water was passed through a pre-combusted (450°C, 4 hours) glass fibre
687 filter (0.5 µm, Macherey-Nagel, Germany), and the filtrate dispatched into three 40 mL pre-
688 combusted amber borosilicate vials and one 500 mL pre-combusted amber borosilicate bottle.
689 Once on land, we acidified the samples to a pH of 2 by adding HCl. We kept the filtrates in the
690 dark, and either on ice or at ca. 4°C for up to 15 days before shipping them to Cambridge, UK.
691 In the laboratory, we stored the samples at 4°C for up to a week and determined the DOC in
692 each 40 mL vial using a Shimadzu TOC-TNM-L analyser (Shimadzu, Japan). In line with
693 regulatory standards[79], we averaged DOC concentrations from the three replicates when the
694 standard deviation was within 10% of the mean and averaged the two closest values otherwise.
695 We estimated chemodiversity as described for the laboratory experiment but solid-phase
696 extracted 500 mL of water instead of 80 mL and retrieved a total of 8456 formulae. We also
697 evaporated 10 µL of extract and redissolved it in 10 mL of Milli-Q water acidified to a pH of
698 2 using HCl to determine the DOC concentration in the extract and corresponding extraction
699 efficiency (mean = 62%, SD =15%).
700
701 To determine if chemodiversity promoted multifunctionality because it could undergo more
702 biochemical transformations, we also estimated the number of times that each molecular
703 formula could be biochemically transformed using an established approach[19]. Fudyma *et al.*[19]
704 compiled 1255 pairs of molecular formulae associated with commonly observed biotic and
705 abiotic transformations, such as enzymatic reactions that add or subtract CH_2 groups and

706 physical condensation or dehydration, respectively. We considered such a transformation to
707 have occurred if we detected each pair of formulae from a reaction in a sample.
708
709 We estimated the complexity of each compound using an aromaticity index (AI_{mod})[51]. A
710 molecule is considered more complex if it possesses more three-dimensional structures and
711 aromatic rings, which are both characterised by the presence of double bonds. However, atoms
712 such as nitrogen can be linked to carbon with double bonds without necessarily forming
713 aromatic rings. By using AI_{mod}, we considered both the number of double bound equivalents
714 (DBE), which approximates the saturation of each molecular formula, and the number of
715 heteroatoms, which provides insight into how carbon is bound. The number of DBE and AI_{mod}
716 were computed within the ICBM-OCEAN pipeline using equations described by Koch &
717 Dittmar[51,71].
718
719 *Biodiversity sampling*
720 We estimated biodiversity by sampling environmental DNA in 82 of the study lakes. During
721 sampling, we passed 1L of water through a 0.2 µm filter cartridge (Sterivex, Millipore, USA).
722 The cartridges were stored at -20°C until DNA was using an established protocol[80]. Briefly,
723 filter cartridges were opened in a sterile environment, and we placed the filters in a cryotube
724 containing silica and zirconia beads (3.0, 0.7, and 0.1 mm diameter). We then vortexed the
725 tubes at 2,850 rpm for 15 min to detach cells from the filter. To ensure DNA was released into
726 solution from lysed cells, we added 0.6 mL of phenol-chloroform-isopropanol (25:24:1), 0.6
727 mL of 5% cetrimonium bromide, 60 µL of 10% sodium dodecyl sulfate, and 60 µL 10% N-
728 lauroylsarcosine before vortexing the tubes a second time at 2,850 rpm for 15 min. We then
729 centrifuged the tubes at 16,000 g for 15 min at 4°C and collected the supernatant. To remove
730 phenol residues, we added 0.6 mL of chloroform-isopropanol (24:1) to the supernatant, gently

731 mixed the samples by inversion, and centrifuged at 16,000 g for 10 min at 4°C. We again
732 collected the supernatant (ca. 0.6 mL) and added 1.2 mL of polyethylene glycol with 1.6M
733 sodium chlorine to precipitate the DNA overnight at 4°C. The following morning, we
734 centrifuged the samples at 17,000 g for 90 min at 4°C to obtain a DNA pellet that we washed
735 with ice-cold (-20°C) 70% ethanol. The pellet was dried at 37°C for 10 min before being
736 dissolved in 50 µL of ultrapure water (Qiagen, Germany). We used the same protocol to extract
737 DNA from a ZymoBIOMICS™ Microbial Community Standard II (Log Distribution) (Zymo
738 Research, USA) and ultrapure water (Qiagen, Germany) that served as positive and negative
739 controls, respectively. The positive control indicated no external contamination, and no DNA
740 was retrieved from the negative control.
741
742 We prepared shotgun metagenomics sequencing libraries with the NEBNext Ultra II FS kit
743 (New England Biolabs, USA) following the manufacturer's instructions. Briefly, DNA was
744 digested for 15 minutes to obtain end polished and A-tailed fragments of ~350bp. Then, we
745 attached Illumina adaptors by incubating the fragments at 37°C for 15 min in presence of a
746 ligation enhancer. We cleaned the samples using AMPure XP beads (Beckman Coulter, USA)
747 and attached Illumina-specific indexes (New England Biosciences, USA). We purified the
748 product using the AMPure XP beads[81], quantified DNA concentration on a Qubit fluorometer
749 (ThermoFisher, USA), and pooled libraries at equimolar concentrations before sequencing on
750 a NovaSeq 6000 PE 250 bp (Illumina, USA) at Novogene (Cambridge, UK), who performed
751 demultiplexing and checked read quality (Q30 ≥ 75% and Q20 = 100%). We obtained a similar
752 coverage for all samples (1st and 3rd quartiles: 4.35×10^7 and 6.70×10^7 reads). We assigned
753 taxonomy to the reads using Kraken2[82] and a database containing bacteria, eukaryotes
754 (including fungi), and archaea. We used Bayesian Reestimation of Abundance with Kraken to

755 obtain read counts at the genus level[83] and estimated biodiversity as the exponential of the
756 Shannon index as for chemodiversity.
757
758 *Ecosystem functioning in the field survey*
759 At each field site, we measured three functions related to ecosystem metabolism (maximum
760 primary production, total respiration, and bacterial protein production) and two functions
761 related to carbon cycling (CH_4 and CO_2 emissions) from the lake surface.
762
763 Primary production and respiration were obtained based on differences in dissolved O_2 as
764 previously described[84]. First, we filled six glass serum bottles with 20 mL of lake water. All
765 bottles were crimped gastight without air bubbles. Two bottles were processed directly, two
766 bottles were incubated in the dark for 24 hours, and the last two bottles faced a cold light
767 (5000K) lamp and received ca. 34,000 lux for 3 hours. For each bottle, we measured the
768 dissolved O_2 concentration using optical fibre sensors (Presens, Germany). We recorded
769 dissolved O_2 concentration every second for approximately five minutes until we reached a
770 plateau. Then, we averaged the last 30 values of the plateau to infer the dissolved O_2
771 concentration inside the serum bottles. Primary production was estimated as the difference in
772 dissolved O_2 between the start of the incubation and after exposure to light. Respiration was
773 the difference in dissolved O_2 between the start and after 24 hours of dark incubation. We
774 averaged duplicates to obtain one value per lake.
775
776 Bacteria protein production (BPP) was estimated based on ^{3}H-leucine incorporation[85]. Directly
777 after sampling, we filled six tubes (Eppendorf, Germany) containing 17nM of ^{3}H-leucine with
778 1.5 mL of lake water. We previously added 300 µL of trichloroacetic acid to three tubes that
779 served as blanks. All tubes were immersed inside the lake in a dark container at ca. 0.5 m depth

780 for exactly 1 hour. We terminated the incubation by adding 300 µL of trichloroacetic acid and
781 preserved the samples on ice until we reached the lake shore – the journey time varied from 1
782 to 30 mins. Then, we centrifuged the samples for 10 min at 16,000 g, removed the supernatant,
783 washed the pellets with 1 mL of 5% trichloroacetic acid, and centrifuged again for 10 min at
784 16,000 g. We removed the supernatant and let the pellets air-dry for up to 10 min. We added 1
785 mL of scintillation cocktail (HiSafe, Perkin Elmer, USA) and measured counts per minutes
786 with a portable liquid scintillation counter (Triathler, Hidex, Finland). To obtain the final BPP,
787 we converted counts per minutes into degradation per minutes, subtracted the value of the mean
788 of the corresponding blanks and expressed the results in mg C L^{-1} day^{-1} as previously
789 described[86]. We averaged the values from the three replicates when the standard deviation was
790 within 10% of the mean and averaged the two closest values otherwise.
791
792 CH_4 and CO_2 emissions were measured using the floating chamber technique[87]. We used a
793 chamber with a 0.07 m^2 perimeter and 0.002 m^3 volume that was covered in aluminium tape to
794 prevent heating. We connected the chamber to an ultraportable greenhouse gas analyser (Los
795 Gatos Research, USA) that uses off-axis integrated cavity output spectroscopy using gastight
796 PVC tubing. Before starting any measurements, we let the chamber equilibrate with the
797 atmosphere by waiting until the CH_4 and CO_2 curves plateaued. Then, we placed the chamber
798 on the water with 3 cm of the chamber immersed underwater. We monitored the concentration
799 of CH_4 and CO_2 inside the chamber every second for 15 mins. We estimated fluxes as the linear
800 slope between the gas concentration and time[87]. We repeated the process (atmospheric
801 equilibration and measurement) twice, and a third time if the two fluxes differed by more than
802 10%. We then averaged the slopes to estimate CH_4 and CO_2 fluxes.
803

804 To calculate a multifunctionality index that reflects human values of nature, we assessed the
805 societal benefit of each measured function[1]. We considered that society benefits from higher
806 primary production because it can fuel food web productivity and fix atmospheric carbon into
807 biomass. Similarly, larger BPP benefits society because it increases the amount of carbon in
808 the food web and can store carbon for longer inside lakes. However, higher greenhouse gas
809 (CO_2 and CH_4) emissions negatively impact society due to their effect on climate. Higher
810 microbial respiration similarly consumes the carbon stored in lakes and emits CO_2 into the
811 environment, some of which will escape to the atmosphere. We therefore multiplied each of
812 CO_2 emissions, CH_4 emissions and respiration rate by -1 to ensure that they had the same
813 directionality of their societal benefits as primary production and BPP, that is, all functions
814 benefited society more as they increased in value. This decision seemingly contrasted with the
815 lab experiment where we calculated absolute values because both increases and decreases in
816 DOC and TDN indicate greater rates of beneficial biogeochemical cycling. However, both
817 approaches are consistent with the idea that different functions are valued in different ways.
818 We then calculated multifunctionality as for the laboratory experiment by summing *z*-scores of
819 each function.
820
821 *Predictors of chemodiversity and ecosystem function*
822 Ecosystem functions and chemodiversity are often influenced by physicochemical
823 variables[16,36] and climate[8,9]. Therefore, during the field study, we recorded water temperature
824 and pH using a multiprobe (HI-99171, Hanna Instruments, USA), measured water colour as
825 the absorbance at 254nm using a spectrophotometer (Flame UV-Vis, Ocean Optics, USA), and
826 measured chlorophyll *a* using a portable fluorometer (FluoroProbe III, bbe Moldaenke,
827 Germany). In the lab, we measured TDN concentrations concurrent with DOC concentrations
828 on the Shimadzu TOC-TNM-L analyser. Climate variables were the mean annual temperature

829 (MAT) and mean annual precipitation (MAP) of each lake at a ca. 31km resolution compiled
830 by the European Centre for Medium-Range Weather Forecasts ERA5 data product[88]. We
831 clipped the ERA5 analysis to the area of each catchment retrieved from the Global Lake area,
832 Climate, and Population dataset[89] and averaged across months for all pixels in a catchment to
833 obtain MAT and MAP. Because ecosystem function can also depend on the surrounding
834 catchment, a net primary production (NPP) raster (500m resolution) based on MODIS satellite
835 images[90] was clipped to each catchment and averaged across all pixels in a catchment to derive
836 NPP.
837
838 *Testing the effects of chemodiversity and biodiversity on ecosystem multifunctionality*
839 We first determined which variables influenced multifunctionality in the lab experiment and
840 field survey by fitting a linear model that only contained biodiversity and chemodiversity. We
841 then fitted a second linear model that included an interaction between biodiversity and
842 chemodiversity to test if any effects of chemodiversity on multifunctionality were conditional
843 on biodiversity. For the lab experiment, chemodiversity and biodiversity were treated as
844 factors, as we only had ≤4 levels of each variable, whereas we used them as continuous
845 variables in the field survey. We used the same model for testing the effect of chemodiversity
846 on multifunctionality as for testing the effect of chemodiversity on each individual function in
847 the laboratory experiment. However, we added mean annual temperature, water colour, lake
848 area, and chlorophyll *a* concentration to model each function in the field survey because
849 previous studies have revealed that these variables also predict ecosystem function[23,24]. All
850 models were fitted with R version 4.1.1[78].
851
852 In addition to the linear models, we tested explicit hypotheses about the drivers of
853 chemodiversity-ecosystem multifunctionality relationships using a path analysis. Here, we

854 designed a path analysis explaining variation in multifunctionality, chemodiversity, the number
855 of biochemical transformations, and DOM complexity. We hypothesized that variation in
856 multifunctionality was explained by chemodiversity, as well as biodiversity, MAT, and MAP,
857 which all predict ecosystem multifunctionality in other studies[3,10,38]. We then hypothesized that
858 chemodiversity was explained by its complexity and potential transformation consistent with
859 established theories for the reactivity of DOM[27]. Based on the same theories, we hypothesized
860 that the number of times a molecule was transformed was also influenced by its complexity
861 and all the environmental variables we measured, including biodiversity. Finally, we
862 hypothesized that DOM complexity was influenced by terrestrial primary production that
863 naturally produces many complex compounds that are leached into downstream waters[65].
864 Although microbial activity also increases complexity[13], we expected its effect to be minimal
865 because most DOM in lakes originates from terrestrial production[16]. We then fitted this path
866 analysis to lake-level observations using the piecewiseSEM package[91] in R, which translates
867 our hypotheses into a set of linear models. After fitting the initial path analysis, we included
868 missing paths identified by tests of directed separation as described in Fig. S13.
869
870 *Identifying the production and consumption of individual compounds*
871 For the laboratory experiment, we first identified individual molecular formulae that were
872 consumed or produced by microbes. We compared the relative abundance of each molecular
873 formula at the end of the incubation experiment from all biodiversity levels against the relative
874 abundance of the same formula in the control with a Welch's t-test. We only considered the
875 final time point of the control to account for compounds that could be produced or consumed
876 simply due to abiotic processes. We considered compounds were produced and consumed if
877 their abundance significantly increased and decreased compared to the control, respectively.

878 We then summed the total number of produced and consumed compounds at each
879 chemodiversity level.
880
881 For the field survey, we tested how the consumption and production of formulae was associated
882 with each function that varied with chemodiversity across Europe (CH_4 and CO_2 emissions and
883 bacterial biomass production). We correlated the relative abundance of each formula in each
884 lake with the value of each function in the corresponding lake using Spearman rank
885 correlations. We considered statistically significant negative and positive correlations to
886 indicate that consumption and production of a formula, respectively, was associated with a
887 given function.
888
889 *Extrapolating chemodiversity and ecosystem function from our field survey*
890 To extrapolate chemodiversity and each measured function that varied with chemodiversity
891 across Europe (CH_4 and CO_2 emissions and bacterial biomass production), we first fitted linear
892 models to these variables. The models included all physio-chemical and environmental
893 variables measured in the field survey, and we included chemodiversity and biodiversity as
894 predictors when modelling each function. Then, we simplified the full models with stepwise
895 variable selection guided by the Akaike information criterion (AIC). We sequentially removed
896 the least supported variable, that is, most increased AIC when retained. If all variables reduced
897 AIC when retained, we dropped the variable that least reduced AIC by <2 units[92]. We repeated
898 this selection process until all variables retained in the model reduced AIC by ≥2 units. The
899 goal of this analysis was to predict each function as accurately as possible to extrapolate across
900 Europe rather than simply compare the importance of chemodiversity versus biodiversity as
901 when modelling multifunctionality. Model selection indicated that chemodiversity was best

902 explained by NPP, MAP, and chlorophyll *a* concentration (Table S3), which was identical to
903 the variables that explained variation in chemodiversity within our path analysis (Figs 2, S13).
904 To extrapolate our linear models across Europe in 2019, we forced them with the same
905 MAP and NPP rasters described above. The other predictors required in our models were water
906 colour and chlorophyll *a* concentration (Table S3, S4), which are unavailable at a European-
907 scale in publicly available datasets. Therefore, we predicted water colour from elevation, MAP,
908 and lake area using a linear model identical to past global studies predicting water colour[75].
909 Similarly, we predicted chlorophyll *a* from MAT, elevation, MAP, and lake area using another
910 linear model. Past global studies have used all these same variables to predict chlorophyll *a*,
911 except that they included TP, which was also unavailable from publicly available datasets at a
912 European-scale[23,24]. Although the models explained 25% and 29% of the variation in water
913 colour and chlorophyll *a*, respectively, the errors introduced by using these models are not
914 relevant for our calculations focusing on relative changes over time (see section below).
915
916 *Forecasting the future functioning of lakes*
917 We estimated how each function that varied with chemodiversity (CH_4 and CO_2 emissions and
918 bacterial biomass production) would change across Europe by 2100 solely because of changes
919 to chemodiversity. To predict chemodiversity in 2100, we first obtained the IPSL-CM6A[60],
920 HadGEM3-GC31[93], and ACCESS-ESM1.5[94] climate forecasts for 2100. We selected these
921 forecasts because they emphasise the relevance of terrestrial-atmosphere exchange and are
922 frequently used in climate studies[95]. We downloaded rasters of MAT and MAP for likely
923 (SSP2) and worst case (SSP5) scenarios for future changes in atmospheric CO_2 concentrations
924 for each forecast from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/).
925 NPP predictions for each forecast were downloaded from the Earth System Grid Federation
926 servers (https://esgf-node.llnl.gov/search/cmip6/). Because no NPP data were available for the

927 HadGEM3-GC31 model, we used the average of the IPSL-CM6A and ACCESS-ESM1.5
928 models. We then input these rasters into our linear model of chemodiversity to predict
929 chemodiversity in 2100 across Europe (Table S3). To generate the predictor of chlorophyll *a*,
930 we again used the model described in the previous section but with MAT and MAP from the
931 climate forecasts, as well as elevation and lake area. The chemodiversity raster was then input
932 into the linear models of CH_4 and CO_2 emissions and bacterial biomass production (Table S4).
933 All predictors other than chemodiversity were set at their 2019 values. Rather than focus on
934 absolute values of each function, we calculated the change in each function between 2019 and
935 2100 when only changing chemodiversity, that is, holding all other variables constant. Because
936 errors associated with models of water colour and chlorophyll *a* are likely to be systematic, that
937 is, non-random, the errors should remain similar between time periods with chemodiversity
938 being the sole difference in the comparison. Furthermore, estimates of the magnitude of
939 changes of CH_4 and CO_2 for all of Europe were well in line with previous global studies[23,24],
940 further suggesting that errors in absolute values of water colour and chlorophyll *a* were of
941 minor importance.
942
943 **References**
944 1. Manning, P. *et al.* Redefining ecosystem multifunctionality. *Nature Ecology and*
945 *Evolution* **2**, 427–436 (2018).
946 2. Naeem, S., Duffy, J. E. & Zavaleta, E. The Functions of Biological Diversity in an Age
947 of Extinction. *Science* **336**, 1401–1406 (2012).
948 3. Delgado-Baquerizo, M. *et al.* Microbial diversity drives multifunctionality in terrestrial
949 ecosystems. *Nature Communications* **7**, 10541 (2016).
950 4. Guillemette, F., Leigh Mccallister, S. & Del Giorgio, P. A. Selective consumption and
951 metabolic allocation of terrestrial and algal carbon determine allochthony in lake bacteria.
952 *ISME Journal* **10**, (2016).

953 5. Tanentzap, A. J. *et al.* Terrestrial support of lake food webs: Synthesis reveals controls
954 over cross-ecosystem resource use. *Science Advances* **3**, (2017).
955 6. Tanentzap, A. J. *et al.* Chemical and microbial diversity covary in fresh water to
956 influence ecosystem functioning. *Proceedings of the National Academy of Sciences of the*
957 *United States of America* **116**, 24689–24695 (2019).
958 7. Bruelheide, H. *et al.* Global trait–environment relationships of plant communities. *Nature*
959 *Ecology and Evolution* **2**, 1906–1917 (2018).
960 8. Maestre, F. T. *et al.* Plant species richness and ecosystem multifunctionality in global
961 drylands. *Science* **335**, 214–218 (2012).
962 9. Lefcheck, J. S. *et al.* Biodiversity enhances ecosystem multifunctionality across trophic
963 levels and habitats. *Nature Communications* **6**, 6936 (2015).
964 10. Jing, X. *et al.* The links between ecosystem multifunctionality and above- and
965 belowground biodiversity are mediated by climate. *Nat Commun* **6**, 8159 (2015).
966 11. Noriega-Ortega, B. E. *et al.* Does the chemodiversity of bacterial exometabolomes
967 sustain the chemodiversity of marine dissolved organic matter? *Frontiers in*
968 *Microbiology* **10**, 215 (2019).
969 12. Graham, E. B. *et al.* Multi ’omics comparison reveals metabolome biochemistry, not
970 microbiome composition or gene expression, corresponds to elevated biogeochemical
971 function in the hyporheic zone. *Science of the Total Environment* **642**, 742–753 (2018).
972 13. Dittmar, T. *et al.* Enigmatic persistence of dissolved organic matter in the ocean. *Nature*
973 *Reviews Earth and Environment* **2**, 570–583 (2021).
974 14. Tanentzap, A. J. & Fonvielle, J. A. Chemodiversity in freshwater health. *Science* **383**,
975 1412–1414 (2024).

15. Testa, B., Vistoli, G., Pedretti, A. & Bojarski, A. J. Atomic Diversity, Molecular Diversity, and Chemical Diversity: The Concept of Chemodiversity. *Chemistry & Biodiversity* **6**, 1145–1151 (2009).
16. Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nature Communications* **5**, 3804 (2014).
17. Hall, S. J., Ye, C., Weintraub, S. R. & Hockaday, W. C. Molecular trade-offs in soil organic carbon composition at continental scale. *Nature Geoscience* **13**, 687–692 (2020).
18. Osterholz, H., Niggemann, J., Giebel, H. A., Simon, M. & Dittmar, T. Inefficient microbial production of refractory dissolved organic matter in the ocean. *Nature Communications* **6**, 1–8 (2015).
19. Fudyma, J. D., Chu, R. K., Graf Grachet, N., Stegen, J. C. & Tfaily, M. M. Coupled Biotic-Abiotic Processes Control Biogeochemical Cycling of Dissolved Organic Matter in the Columbia River Hyporheic Zone. *Frontiers in Water* **2**, (2021).
20. Helms, J. R. *et al.* Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography* **53**, 955–969 (2008).
21. Rose, K. C., Winslow, L. A., Read, J. S. & Hansen, G. J. A. Climate-induced warming of lakes can be either amplified or suppressed by trends in water clarity. *Limnology and Oceanography Letters* **1**, 44–53 (2016).
22. Senar, O. E., Creed, I. F. & Trick, C. G. Lake browning may fuel phytoplankton biomass and trigger shifts in phytoplankton communities in temperate lakes. *Aquat Sci* **83**, 21 (2021).
23. Rosentreter, J. A. *et al.* Half of global methane emissions come from highly variable aquatic ecosystem sources. *Nature Geoscience* **14**, 225–230 (2021).

1001 24. DelSontro, T., Beaulieu, J. J. & Downing, J. A. Greenhouse gas emissions from lakes and
1002 impoundments: Upscaling in the face of global change. *Limnology And Oceanography*
1003 *Letters* **3**, 64–75 (2018).
1004 25. Benner, R. & Amon, R. M. W. The Size-Reactivity Continuum of Major Bioelements in
1005 the Ocean. *Annual Review of Marine Science* **7**, 185–205 (2015).
1006 26. Hutchins, R. H. S. *et al.* The Optical, Chemical, and Molecular Dissolved Organic Matter
1007 Succession Along a Boreal Soil-Stream-River Continuum. *Journal of Geophysical*
1008 *Research: Biogeosciences* **122**, 2892–2908 (2017).
1009 27. Kothawala, D. N., Kellerman, A. M., Catalán, N. & Tranvik, L. J. Organic Matter
1010 Degradation across Ecosystem Boundaries: The Need for a Unified Conceptualization.
1011 *Trends in Ecology & Evolution* **36**, 113–122 (2020).
1012 28. Lehner, B. & Döll, P. Development and validation of a global database of lakes,
1013 reservoirs and wetlands. *Journal of Hydrology* **296**, 1–22 (2004).
1014 29. Mentges, A., Feenders, C., Seibt, M., Blasius, B. & Dittmar, T. Functional Molecular
1015 Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Frontiers*
1016 *in Marine Science* **4**, 194 (2017).
1017 30. Kritzberg, E. S. *et al.* Browning of freshwaters: Consequences to ecosystem services,
1018 underlying drivers, and potential mitigation measures. *Ambio* **49**, 375–390 (2020).
1019 31. Kolmakova, O. V. *et al.* Effects of zooplankton carcasses degradation on freshwater
1020 bacterial community composition and implications for carbon cycling. *Environmental*
1021 *Microbiology* **21**, 34–49 (2019).
1022 32. Gamfeldt, L. & Roger, F. Revisiting the biodiversity-ecosystem multifunctionality
1023 relationship. *Nature Ecology and Evolution* **1**, (2017).

1024 33. Le Bagousse-Pinguet, Y. *et al.* Phylogenetic, functional, and taxonomic richness have
1025 both positive and negative effects on ecosystem multifunctionality. *Proceedings of the*
1026 *National Academy of Sciences* **116**, 8419–8424 (2019).
1027 34. Foster, K. R. & Bell, T. Competition, Not Cooperation, Dominates Interactions among
1028 Culturable Microbial Species. *Current Biology* **22**, 1845–1850 (2012).
1029 35. Rivett, D. W. & Bell, T. Abundance determines the functional role of bacterial
1030 phylotypes in complex communities. *Nat Microbiol* **3**, 767–772 (2018).
1031 36. Delgado-Baquerizo, M. *et al.* Multiple elements of soil biodiversity drive ecosystem
1032 functions across biomes. *Nat Ecol Evol* **4**, 210–220 (2020).
1033 37. Byrnes, J. E. K. *et al.* Investigating the relationship between biodiversity and ecosystem
1034 multifunctionality: Challenges and solutions. *Methods in Ecology and Evolution* **5**, 111–
1035 124 (2014).
1036 38. Allan, E. *et al.* Land use intensification alters ecosystem multifunctionality via loss of
1037 biodiversity and changes to functional composition. *Ecology Letters* (2015)
1038 doi:10.1111/ele.12469.
1039 39. Kellerman, A. M. *et al.* Unifying Concepts Linking Dissolved Organic Matter
1040 Composition to Persistence in Aquatic Ecosystems. *Environmental Science and*
1041 *Technology* **52**, 2538–2548 (2018).
1042 40. Merbach, W., Ruppel, S. & Schulze, J. Dinitrogen Fixation of Microbe-Plant
1043 Associations as Affected by Nitrate and Ammonium Supply. *Isotopes in Environmental*
1044 *and Health Studies* **34**, (1998).
1045 41. Jung, H. *et al.* Comparative genomic analysis of Chryseobacterium species: deep insights
1046 into plant-growth-promoting and halotolerant capacities. *Microbial Genomics* **9**, 001108
1047 (2023).

42. Kraemer, S. A. *et al.* A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. *ISME Journal* **14**, 3011–3023 (2020).
43. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
44. 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 Communications in Mass Spectrometry* **29**, 2385–2401 (2015).
45. Günthel, M. *et al.* Contribution of oxic methane production to surface methane emission in lakes and its global importance. *Nat Commun* **10**, 5497 (2019).
46. Ernst, L. *et al.* Methane formation driven by reactive oxygen species across all living organisms. *Nature* **603**, 482–487 (2022).
47. Mosher, J. J., Kaplan, L. A., Podgorski, D. C., McKenna, A. M. & Marshall, A. G. Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. *Biogeochemistry* **124**, 371–385 (2015).
48. Lapierre, J.-F., Seekell, D. A. & del Giorgio, P. A. Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. *Global Change Biology* **21**, 4425–4435 (2015).
49. Fuchs, G., Boll, M. & Heider, J. Microbial degradation of aromatic compounds — from one strategy to four. *Nat Rev Microbiol* **9**, 803–816 (2011).
50. Zhang, R., Wang, J., Milligan, S. & Yan, Y. Microbial utilization of lignin-derived aromatics: Via a synthetic catechol meta-cleavage pathway. *Green Chemistry* **23**, 8238–8250 (2021).

51. Koch, B. P. & Dittmar, T. From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry* **20**, 926–932 (2006).
52. Arrieta, J. M. *et al.* Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* **348**, 331–333 (2015).
53. Weng, J. K., Philippe, R. N. & Noel, J. P. The rise of chemodiversity in plants. *Science* **336**, 1667–1670 (2012).
54. Bertilsson, S. & Tranvik, L. J. Photochemical transformation of dissolved organic matter in lakes. *Limnology and Oceanography* **45**, 753–762 (2000).
55. Olson, C. R., Solomon, C. T. & Jones, S. E. Shifting limitation of primary production: experimental support for a new model in lake ecosystems. *Ecology Letters* **23**, 1800–1808 (2020).
56. Dittmar, T., Hertkorn, N., Kattner, G. & Lara, R. J. Mangroves, a major source of dissolved organic carbon to the oceans. *Global Biogeochemical Cycles* **20**, (2006).
57. Cuhel, R. L. & Aguilar, C. Ecosystem Transformations of the Laurentian Great Lake Michigan by Nonindigenous Biological Invaders. *Annual Review of Marine Science* **5**, 289–320 (2013).
58. Waska, H., Koschinsky, A., Ruiz Chancho, M. J. & Dittmar, T. Investigating the potential of solid-phase extraction and Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) for the isolation and identification of dissolved metal–organic complexes from natural waters. *Marine Chemistry* **173**, 78–92 (2015).
59. Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5**, 782–791 (2007).
60. Boucher, O. *et al.* Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems* **12**, e2019MS002010 (2020).

61. Zhang, W., Shen, J. & Wang, J. Linking pollution to biodiversity and ecosystem multifunctionality across benthic-pelagic habitats of a large eutrophic lake: A whole-ecosystem perspective. *Environmental Pollution* **285**, 117501 (2021).

62. Zhuang, Q. *et al.* Current and Future Global Lake Methane Emissions: A Process-Based Modeling Analysis. *Journal of Geophysical Research: Biogeosciences* **128**, e2022JG007137 (2023).

63. Messager, M. L., Lehner, B., Grill, G., Nedeva, I. & Schmitt, O. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications* **7**, 13603 (2016).

64. O'Neill, B. C. *et al.* The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development* **9**, 3461–3482 (2016).

65. Defossez, E. *et al.* Spatial and evolutionary predictability of phytochemical diversity. *Proceedings of the National Academy of Sciences of the United States of America* **118**, (2021).

66. Díaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).

67. Brienen, R. J. *et al.* Forest carbon sink neutralized by pervasive growth-lifespan trade-offs. *Nature communications* **11**, 4241 (2020).

68. Scheuerl, T. *et al.* Bacterial adaptation is constrained in complex communities. *Nat Commun* **11**, 754 (2020).

69. Gheorghiu, C., Smith, D. S., Al-Reasi, H. A., McGeer, J. C. & Wilkie, M. P. Influence of natural organic matter (NOM) quality on Cu–gill binding in the rainbow trout (Oncorhynchus mykiss). *Aquatic Toxicology* **97**, 343–352 (2010).

70. 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. *Limnology and Oceanography: Methods* **6**, 230–235 (2008).

71. Merder, J. *et al.* ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Analytical Chemistry* **92**, 6832–6838 (2020).

72. Shannon, C. E. C. & Weaver, W. *The Mathematical Theory of Information*. *Urbana University of Illinois Press* (1949). doi:10.1007/978-1-4615-0975-2.

73. Jost, L. Entropy and diversity. *Oikos* **113**, 363–375 (2006).

74. European Environmental Agency. WISE Water Framework Directive Database. https://www.eea.europa.eu/data-and-maps/data/wise-wfd-4 (2019).

75. Toming, K. *et al.* Predicting lake dissolved organic carbon at a global scale. *Scientific Reports* **10**, 8471 (2020).

76. McMaster, G. S. & Wilhelm, W. W. Growing degree-days: One equation, two interpretations. *Agricultural and Forest Meteorology* **87**, 291–300 (1997).

77. Scott Chamberlain. rnoaa: 'NOAA' Weather Data from R. R package version 0.8.4. *https://CRAN.R-project.org/package=rnoaa* (2019).

78. R Core team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Preprint at https://doi.org/10.2788/95827 (2017).

79. Potter, B. B. & Wimsatt, J. C. USEPA method 415.3: Quantifying TOC, DOC, and SUVA. *Journal AWWA* **104**, E358–E369 (2012).

80. Nercessian, O., Noyes, E., Kalyuzhnaya, M. G., Lidstrom, M. E. & Chistoserdova, L. Bacterial populations active in metabolism of C1 compounds in the sediment of Lake Washington, a freshwater lake. *Applied and Environmental Microbiology* **71**, 6885–6899 (2005).

81. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

82. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 1–13 (2019).

83. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science* **2017**, e104 (2017).

84. Revsbech, N. P., JØrgensen, B. B. & Brix, O. Primary production of microalgae in sediments measured by oxygen microprofile, H14CO3 - fixation, and oxygen exchange methods. *Limnology and Oceanography* **26**, 717–730 (1981).

85. Simon, M. & Azam, F. Protein content and protein synthesis rates of planktonic marine bacteria. *Marine Ecology Progress Series* **51**, 201–213 (1989).

86. Amand, A. S. *et al.* CASCADING TROPHIC INTERACTIONS PROJECT. (1978).

87. Bastviken, D., Cole, J., Pace, M. & Tranvik, L. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global Biogeochemical Cycles* **18**, GB4009 (2004).

88. Copernicus Climate Change Service. C3S ERA5-Land reanalysis. Copernicus Climate Change Service. (2019) doi:10.24381/cds.68d2bb30.

89. Meyer, M. F., Labou, S. G., Cramer, A. N., Brousil, M. R. & Luff, B. T. The global lake area, climate, and population dataset. *Scientific Data* **7**, 174 (2020).

90. Running, S. & Zhao, M. MOD17A3HGF MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500 m SIN Grid V006. 2019, distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD17A3HGF.006 (2020).

91. Lefcheck, J. S. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution* **7**, 573–579 (2016).

92. Burnham, K. P. & Anderson, D. R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* **33**, 261–304.

93. Williams, K. D. *et al.* The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems* **10**, 357–380 (2018).

94. Bi, D. *et al.* Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. *JSHESS* **70**, 225 (2020).

95. Friedlingstein, P. *et al.* Global Carbon Budget 2021. *Earth System Science Data* **14**, 1917–2005 (2022).