Niche separation increases with genetic distance among bloom-forming cyanobacteria

Authors

Nicolas Tromas[1*#], Zofia E. Taranu[2*], Bryan D Martin[3], Amy Willis[4], Nathalie Fortin[5], Charles W. Greer[5] and B. Jesse Shapiro[1#]

Author affiliations

1- Département de sciences biologiques, Université de Montréal, 90 Vincent-d'Indy, Montréal, QC, Montréal, QC H2V 2S9, Canada

2- Department of Biology, University of Ottawa, 30 Marie Curie, ON, Ottawa, K1N 6N5, Canada

3- Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195

4- Department of Biostatistics, University of Washington, Box 357232, Seattle, 98195

5- National Research Council Canada, Energy, Mining and Environment, 6100 Royalmount Avenue, Montréal, QC H4P 2R2, Canada

* These authors contributed equally to this work

#Corresponding authors: B. Jesse Shapiro. Phone: 514-343-6033. E-mail: jesse.shapiro@umontreal.ca; Nicolas Tromas. Phone 514-343-3188. E-mail: nicolas.tromas@umontreal.ca.

SUMMARY

Bacterial communities are composed of distinct groups of potentially interacting lineages, each thought to occupy a distinct ecological niche. It remains unclear, however, how quickly niche preference evolves and whether more closely related lineages are more likely to share ecological niches. We addressed these questions by following the dynamics of two bloom-forming cyanobacterial genera over an 8-year time-course in Lake Champlain, Canada, using 16S amplicon sequencing and measurements of several environmental parameters. The two genera, *Microcystis* (*M*) and *Dolichospermum* (*D*), are frequently observed simultaneously during bloom events and thus have partially overlapping niches. However, the extent of their niche overlap is debated, and it is also unclear to what extent niche partitioning occurs among strains within each genus. To identify strains within each genus, we applied minimum entropy decomposition (MED) to 16S rRNA gene sequences. We confirmed that at a genus level, *M* and *D* have different preferences for nitrogen and phosphorus concentrations. Within each genus, we also identified strains differentially associated with temperature, precipitation, and concentrations of nutrients and toxins. In general, niche similarity between strains (as measured by co-occurrence over time) declined with genetic distance. This pattern is consistent with habitat filtering – in which closely-related taxa are ecologically similar, and therefore tend to co-occur under similar environmental conditions. In contrast with this general pattern, similarity in certain niche dimensions (notably particulate nitrogen and phosphorus) did not decline linearly with genetic distance, and instead showed a complex polynomial relationship. This observation suggests the importance of processes other than habitat filtering – such as competition between closely-related taxa, or convergent trait evolution in distantly-related taxa – in shaping particular traits in microbial communities.

49 **INTRODUCTION**

50 Microbial communities are composed of potentially diverse groups of lineages, which
51 must be sufficiently similar to survive in the same habitat, yet sufficiently dissimilar to occupy
52 distinct ecological niches and avoid competition. This tension between selection for common
53 traits to survive in a common environment (habitat filtering) and selection for divergent traits to
54 reduce competition (niche partitioning among closely-related species) was recognized by Darwin,
55 and the relative impacts of the two processes on communities are still debated (Cavender-Bares *et*
56 *al*., 2009). A pioneering study of microbial communities using phylogenetic marker genes found
57 evidence for phylogenetic clustering, suggesting the importance of habitat filtering in selecting
58 for closely-related taxa sharing specific traits allowing them to survive in a given habitat (Horner-
59 Devine and Bohannan, 2006). However, phylogenetic overdispersion (the opposite pattern as
60 phylogenetic clustering) has also been observed, suggesting that competition between closely-
61 related taxa can lead to niche partitioning (Koeppel and Wu, 2014). Importantly, the power to
62 detect phylogenetic overdispersion depends on the phylogenetic resolution (*e.g.* whether
63 operational taxonomic units are defined at 97, 98 or 99% identity in a marker gene) (Koeppel and
64 Wu, 2014).

65 Beyond searching for phylogenetic patterns of clustering or overdispersion, explicitly
66 considering the associations between microbial traits and niches can help understand the selective
67 pressures shaping microbial communities on different evolutionary time scales. It is known that
68 certain traits (*e.g.* salinity preference, methanogenesis) are relatively slow-evolving and thus
69 restricted to only certain lineages, whereas other traits (*e.g.* phage resistance, organic phosphate
70 utilization) can be acquired by a single point mutation or gene acquisition, thus evolving rapidly
71 in response to ecological selection and competition (Martiny *et al*., 2015). Therefore, habitat

filtering might be stronger for slow-evolving traits, while niche partitioning will be more likely for fast-evolving traits.

In this study, we use Hutchinson's definition of a fundamental niche as the set of abiotic conditions under which an organism can survive and reproduce, and a realized niche as the conditions under which it is actually observed in nature, accounting for both abiotic and biotic (*e.g.* competition, predation, cooperation) interactions (Hutchinson 1957). If two species have identical ecological niches, one should competitively exclude the other (Gause 1934; Tilman 1982) unless competition is weak due to abundant resources. In practice, closely related taxa often compete for space and resources (Cavender-Bares *et al*., 2009), favouring specialization to reduce overlap in niche space. For example, coexisting (sympatric) taxa within the same genus tend to have different realized niches, experiencing different seasonal growth patterns or responding differently to environmental parameters (Gray *et al*., 2004; Jaspers and Overmann, 2004; Hunt *et al*, 2008; Simek *et al*, 2010; Jezberra *et al*, 2011; Neuenschwander *et al*., 2015). Pairs of taxa with similar realized niches can be identified when they co-occur in repeated sampling over space and time, and such co-occurrence networks are readily inferred from deep amplicon sequencing datasets (Friedman and Alm, 2012). Typically, niches are considered as features of species. However, when niches are considered as collections of traits or environmental associations, the Hutchinsonian niche concept can be extended to taxonomic groupings more inclusive than species, even if the biological "reality" of such groups is doubtful. Here, we apply the niche concept to both fine-grained (*i.e.* sub-genus level) and coarse-grained (*i.e.* genus level) taxonomic units. We focus mainly on niche specialization within genera, showing that specialization is extensive and that lumping bacterial diversity at the genus level obscures finer-scale niche preferences.

95 Cyanobacteria are widely and naturally present in freshwater ecosystems, and some
96 lineages form blooms under appropriate conditions of temperature and nutrients (Konopka and
97 Broke, 1978; Harke *et al*, 2016). Several studies have shown that different cyanobacterial genera
98 can co-occur during blooms, thus sharing at least some dimensions of their realized niches (Pearl
99 *et al*, 2001; Yamamoto, 2009). Nitrogen (N) and phosphorus (P) can both be limiting for bloom
100 formation, and different cyanobacterial taxa apparently have different preferences for N and P
101 concentrations (Dolman *et al*., 2012). For example, cyanobacteria capable of N-fixation (such as
102 *Dolichospermum*) are associated with more efficient P-utilization at the community level,
103 suggesting P-limitation when N-fixers are abundant (Andersson *et al*., 2015; Olli *et al*., 2015).
104 Clearly, N and P utilization are ecologically important traits for cyanobacteria, and may be
105 important for niche partitioning among closely-related strains. In the marine cyanobacterium
106 *Prochloroccocus*, P uptake and metabolism appears to be relatively fast-evolving perhaps due to
107 horizontal gene transfer of P-related genes (Coleman *et al*., 2010) while light preferences are
108 slow-evolving, and temperature preferences are intermediate (Martiny *et al*., 2015).
109 We investigated niche partitioning within and between *Microcystis* (*M*) and
110 *Dolichospermum* (*D*), two genera of potentially toxigenic cyanobacteria that bloom nearly every
111 year in a large eutrophic North American lake, Lake Champlain. In a previously described 8-year
112 time-course analysis (2006 to 2013) spanning multiple bloom events, we used 16S amplicon
113 sequencing to broadly survey changes in the lake microbial community over time, generally at
114 the genus level (Tromas *et al*., 2017). Here, we use Minimum Entropy Decomposition of
115 amplicon sequences (Eren *et al*., 2014) to identify sub-genus strains (MED nodes; here used
116 interchangeably with "strains") within each of the two dominant cyanobacterial genera, *M* and *D*,
117 at single nucleotide resolution (*i.e.* each MED node is an exact sequence variant, distinguishable
118 from other variants that differ by at least one nucleotide substitution). MED discards low-entropy

119 nucleotide positions, which effectively filters out sequencing errors at the expense of possibly
120 also removing true polymorphism at very low frequency. Therefore, although MED outputs exact
121 sequences that are actually present in the sample (after denoising), it is possible the MED nodes
122 contain finer-scale genetic variation which could be captured using additional marker genes or
123 whole genome sequencing. A previous study demonstrated that oligotypes (similar to MED
124 nodes) lacked the resolution to distinguish toxic and non-toxic *Microcystis* lineages, but could
125 potentially be informative about other, more phylogenetically conserved niches such as eutrophic
126 (nutrient-rich) vs. oligotrophic (nutrient-poor) lake preferences (Berry *et al*., 2017). Toxin
127 production is thought to be fast-evolving because toxin biosynthesis genes are widely distributed
128 across cyanobacterial genera, suggesting rapid gain and loss. While horizontal gene transfer is a
129 likely explanation, transfer is probably more frequent among closely-related lineages because
130 toxin gene trees are congruent with ribosomal phylogenies of distantly-related cyanobacteria
131 genera (Rantala *et al*., 2004).

132 We used a combination of genetic data (diversity of MED nodes) and matched
133 environmental data (*e.g.* temperature, nutrient concentrations, precipitation) to address three
134 specific questions. First, how similar are the niches of the two dominant cyanobacterial genera *M*
135 and *D*? Second, how similar are the niches of strains within each genus? Third, how does niche
136 similarity change with genetic relatedness? We confirm that *M* and *D* are broadly co-occuring
137 during blooms, but have distinct nutrient preferences. We also identified niche partitioning at the
138 sub-genus level, and observe a general decline in realized niche similarity with genetic distance,
139 consistent with habitat filtering. However, certain niche dimensions (particulate nutrient
140 concentrations) show a complex polynomial relationship with genetic distance, suggesting that a
141 combination of habitat filtering and competitive interactions shapes the evolution of these traits.
142

143

144 **MATERIALS AND METHODS**

145

146 **Sampling, DNA extraction, purification and sequencing**

147 Open-water season samples (April to November) collected over 8 years (2006 to 2013)
148 from the photic zone of Missisquoi Bay at two different sites (littoral and pelagic) of Lake
149 Champlain, Quebec, Canada (45°02'45"N, 73°07'58"W) were filtered and extracted for DNA
150 sequencing as described in Tromas *et al*., (2017).

151

152 **Sequence analysis**

153 A total of 7,949,602 sequences of the 16S rRNA gene V4 region were obtained from 150
154 lake samples, with a median of 41,982 per sample as previously described (Tromas *et al*., 2017).
155 These sequences were processed with the default parameters of the SmileTrain pipeline
156 (https://github.com/almlab/SmileTrain/wiki) that includes reads quality filtering, chimera
157 filtering, and merging using USEARCH (version 7.0.1090, http://www.drive5.com/usearch/,
158 default parameter) (Edgar, 2010), Mothur (version 1.33.3) (Schloss *et al*., 2009), Biopython
159 (version 2.7) and custom scripts. Minimum Entropy Decomposition (MED) was then applied to
160 the filtered and merged reads to partition sequence reads into MED nodes (Eren *et al*., 2014).
161 MED was performed using the following parameters –M noise filter set to 500 resulting in ~7%
162 of reads filtered and 941 MED nodes (Data Sheet 2). Samples with less than 1000 reads were
163 removed, yielding a final dataset of 135 samples. Finally, taxonomy was assigned using the
164 assign_taxonomy.py QIIME script (default parameters), and a combination of GreenGenes and a
165 freshwater-specific database (Freshwater database 2016 August 18 release; Newton *et al*., 2011),
166 using TaxAss (https://github.com/McMahonLab/TaxAss, installation date: September 13th 2016;
167 Rohwer *et al*., 2017). After assignment, nodes that belong to Eukaryotes but still present in the

168 database (Cryptophyta, Streptophyta, Chlorophyta and Stramenopiles orders) were removed,
169 leading to a total of 891 nodes.
170
171 **Diversity analysis**
172 Comparing changes in the diversity of *M* strains to the diversity of *D* strains was
173 performed using betta (Willis *et al*., 2017). betta accounts for strains that are present in the
174 environment but not observed in the samples due to incomplete sampling. The number of
175 unobserved strains is estimated based on the number of strains that are observed and their
176 abundances. The total strain diversity was estimated using breakaway (R package v4)which
177 accounts for ecological interactions between strains (Willis and Bunge, 2015).
178
179 **Conditionally rare taxa analysis**
180 We investigated the temporal dynamics of *M* and *D* nodes by measuring the composition for each
181 genus in conditionally rare taxa. The matrix of node absolute abundances was used as input for
182 the R script CRT_Functions_v1.1.R (Shade *et al*., 2014;
183 https://github.com/ShadeLab/ConditionallyRareTaxa) using the default parameters. Conditionally
184 rare taxa are defined as usually-rare taxa that occasionally become very abundant, without
185 showing rhythmic or seasonal patterns.
186
187 **Node–environment relationships analysis**
188 To investigate node–environment relationships, we used an environmental data matrix that
189 included: particulate phosphorus in µg/ L (PP, the difference between TP and DP), particulate
190 nitrogen in mg/ L (PN, the difference between TN and DN), total dissolved phosphorus in µg/ L
191 (DP), total dissolved nitrogen in mg/ L (DN), 1-week-cumulative precipitation in mm and 1-

192 week-average air temperature in Celsius. Total nutrients were measured directly from collected
193 lake water and the dissolved nutrients were measured in filtered water (Glass microfiber Filters
194 grade GF/F, 0.7 microns). The detailed measurements of each environment variable are described
195 in Tromas *et al* (2017). In this previous study, we showed that these environmental variables,
196 over the 8 years, were not correlated with one another.
197
198 **Response to abiotic factors**
199 To analyze the response of each node to abiotic environmental data, we used a Latent
200 Variable Model (LVM) framework, which combines Generalized Linear Models with Bayesian
201 Markov Chain Monte Carlo (MCMC) methods (boral package in R; Hui 2015, Warton *et al.*
202 2015). LVM is a model-based approach for analyzing multivariate data (*e.g.* numerous taxa
203 within the response matrix) that partitions the different drivers of taxa co-occurrence patterns into
204 two components: the first is a regression component, which models the taxon-specific
205 environmental responses, and the second is a latent variable component, which is used to identify
206 residual patterns of co-occurrence resulting from unmeasured factors and/or biotic interactions
207 (Letten *et al.* 2015). In this study, we used LVMs to examine how taxa co-respond to abiotic
208 gradients, and used these co-responses as a proxy for niche similarity. To do so, we extracted the
209 *environmental correlation matrix* from the regression model, which, for any two taxa,
210 corresponds to the correlation between their fitted values ($x_i\beta_j$). In particular, for each node we
211 calculated the predicted probability of mean abundance for each site on linear and nonlinear (*i.e.*
212 $Y \sim X + X^2$) response scales, providing a vector of fitted values. Correlations among the fitted
213 responses of any two taxa were then calculated based upon these vectors (correlating the vector
214 of taxon A with the vector of taxon B, for instance).

215 We first tested LVMs using a response matrix of two columns, corresponding to the
216 relative abundances of the *D* and *M* genera, relative to the rest of the bacterial community. For
217 these models, *D* and *M* abundances were normalized by the total counts of all bacteria using the
218 centered log ratio to correct for data compositionality (clr-inter genus) (Aitchison, 1986; Paliy
219 and Shankar, 2016) using a a zero-replacement procedure as suggested in Gloor and Reid (2016).
220 To then examine dynamics within these genera, we ran separate LVMs for *D* and *M* nodes (*i.e.*
221 response matrix consisted of the *D* and *M* nodes, respectively). For this second set of LVMs,
222 node abundances were normalized by the total counts of *D* and *M*, respectively, to obtain an
223 intra-genus relative abundance (clr-intra genus) for each node. Environmental variables were
224 standardized to mean zero and unit variance to reduce the correlation between the linear (X) and
225 nonlinear (X^2) fit, which also simplified model interpretation and stabilized MCMC sampling. As
226 most of the *D* nodes were conditionally rare, we included only the *D* nodes that were present in at
227 least 70% of samples, to avoid overfitting based on too little data. To complement this
228 multivariate LVM analysis and help visualize the univariate response of each genus and node to
229 the suite of environmental variables, we also tested linear (degree-1 polynomial) and nonlinear
230 (degree-2 polynomial) relationships between each response and explanatory variable. We used
231 AIC to find the single best-fit model (either linear or nonlinear) for each relationship and plotted
232 the best-fit model.

233 Lastly, to test whether co-responses or niche separations were stronger for more closely or
234 distantly related nodes, we extracted the correlations between fitted responses of any two nodes
235 from the *environmental correlation matrix* of each best-fit LVM and examined the relationship
236 between each significant correlation and the pairwise genetic distances of the two respective taxa
237 (percent identity in the V4 region of the 16S gene). Here, significant positive correlations
238 represent co-response between any two taxa whereas significant negative correlations represent

239 the degree of niche separation. The genetic distance between nodes was measured within each
240 genus *M* and *D* using the software MEGA (version 7.0.18) using the p-distance (the proportion of
241 nucleotide sites at which two sequences differ), calculated by dividing the number of sites with
242 nucleotide differences by the total number of sites compared (excluding sites with gaps). R code
243 to reproduce these analyses and relevant figures is provided in Data Sheet 3.

244 **Co-occurrence analysis**

245 Co-occurrences between *Microcystis* or *Dolichospermum* and other taxa were calculated
246 with SparCC (Friedman and Alm, 2012), with 20 iterations to estimate the median correlation of
247 each pair of MED nodes, and 500 bootstraps to assess the statistical significance. Correlations
248 were then filtered for statistical significance ($P < 0.01$) and correlations with $R > \pm 0.6$, were
249 selected to build networks using Cytoscape (version 3.1.0).

250

251 **RESULTS**
252

253 **Do *Microcystis* and *Dolichospermum* co-occur temporally?**

254 We have previously shown that *M* and *D* were the two most dominant bloom forming
255 cyanobacteria in Lake Champlain's Missisquoi Bay (Tromas *et al*., 2017). Both genera were
256 present every year at both littoral and pelagic sampling sites, between 2006 and 2013, with a
257 relative abundance of at least 15% on average during summer (max=24.8%), with the exception
258 of summer 2007 (0.7%) when no substantial blooms occurred (Figure 1). *D* was generally the
259 most dominant genus except for 2009 and 2010 when *M* dominated (Figure S1). *D* and *M* were
260 both present every year, (Figure S1) suggesting that they share similar physico-temporal niche
261 during a bloom, although other aspects of their niches likely differ. For example, *D* can fix
262 nitrogen but *M* cannot (Hajdu *et al*., 2007).

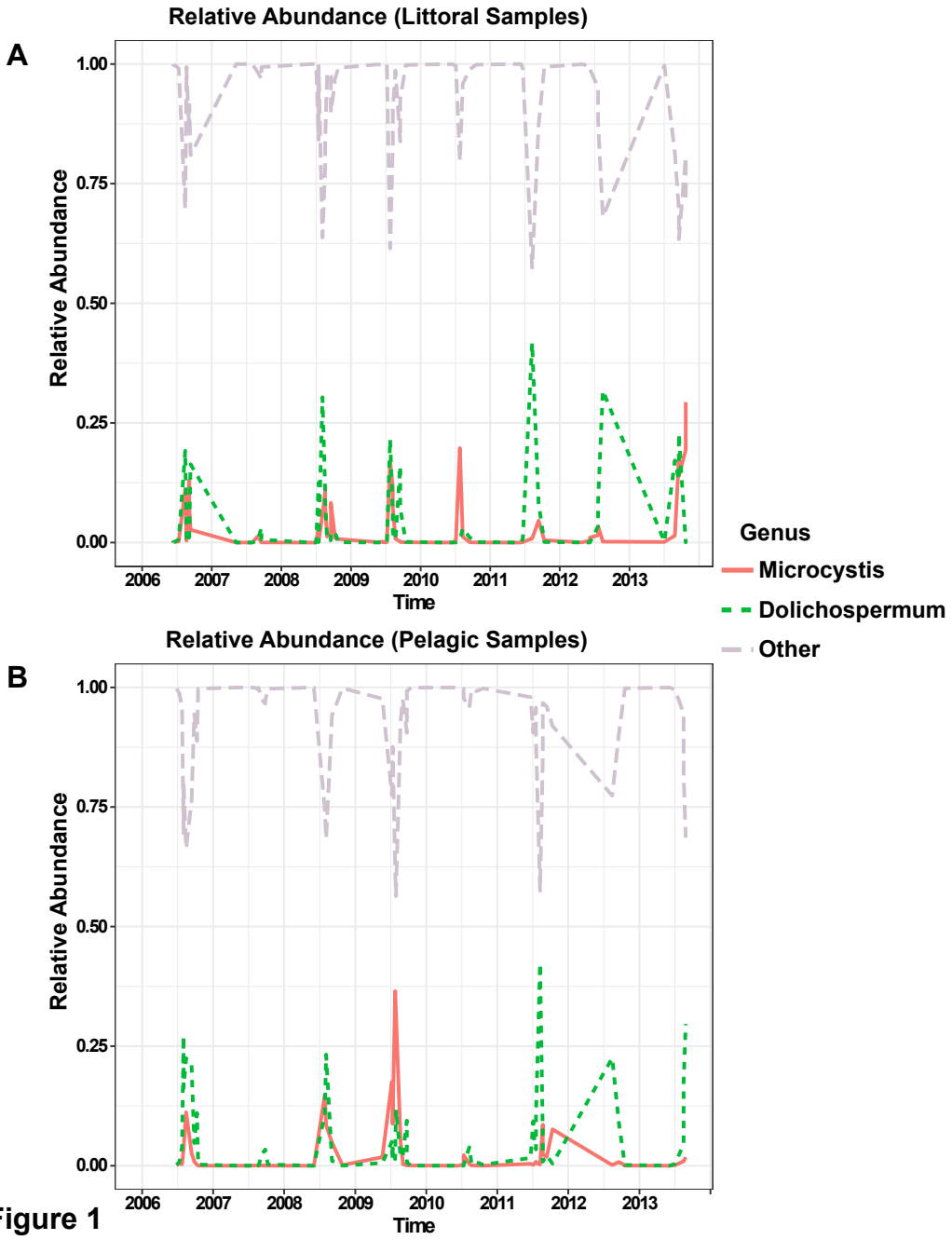


Figure 1. Temporal dynamics of the two dominant cyanobacterial genera over an eight-year time course in littoral (A) and pelagic (B) sampling sites. Relative abundance of *Microcystis* is shown in solid red, *Dolichospermum* in dashed green and the other members of the bacterial community in dashed light violet. The time scale (x-axis) is in units of years.

270 **Are *Microcystis* and *Dolichospermum* equally diverse and dynamic?**

271 To explore the temporal dynamics of the finer-scale taxa within each genus, we partitioned each
272 genus into MED nodes, which we call "strains" (Methods). We obtained 25 *D* strains and 6 *M*
273 strains, and after accounting for low-abundance strains that may be missing from the samples
274 using Breakaway (Methods), we conclude that diversity is significantly lower within *M* than
275 within *D* ($P<0.001$). The difference in the number of nodes is unlikely to be an artefact of
276 sequencing depth, because there is no correlation between the number of sequence reads and the
277 number of nodes per genus (Figure S2). *D* thus appears to be more diverse than *M*.

278 We then examined whether nodes within the *M* and *D* genera followed the same temporal
279 dynamics. We observed that 15/25 *D* nodes were conditionally rare, meaning that they are
280 usually rare but occasionally become relatively abundant, without following any apparent rhythm
281 (Table 1).

282

	Number of conditionally rare nodes	Number of total nodes	Proportion conditional rare / total nodes	Fisher's Exact Test P-value
Whole bacterial community	95	891	0.10	
Dolichospermum	15	25	0.60	$P < 0.001$
Microcystis	1	6	0.16	$P > 0.1$

283

284 **Table 1.** Conditionally rarity analysis. We compared the proportion of conditionally rare *M* or *D*
285 MED nodes to the proportion expected among MED nodes in the entire lake bacterial community
286 using Fisher's exact test.

287

288 The proportion of rare *D* nodes was significantly higher than what is observed in other nodes in
289 the lake community (Fisher's exact test, $P < 0.001$). Several conditionally rare *D* nodes seemed to
290 dominate several times without showing seasonal patterns (Figure S3). Furthermore, we noticed a
291 shift in node composition after 2011, where node D5505 decreased while D2282 increased in
292 relative abundance. In contrast, we only found one conditionally rare *M* node. However, only two
293 *M* nodes (M5732 and M5733) were consistently dominant over time (Figure S4). Similarly, only
294 two *D* nodes, but not always the same two, dominated at any given time (Figure S3). Overall,
295 these results suggest that *D* is more genetically diverse, and that this genetic diversity varies over
296 time. In contrast, *M* is less diverse and more stable over time.
297
298 **Do *Microcystis* and *Dolichospermum* share the same realized niche within the community?**
299 To investigate the niche separation between *D* and *M*, we analyzed their relationships with
300 several environmental conditions measured at the time of sampling (temperature, nutrient
301 concentrations, precipitation and toxin concentrations). We observed that these two
302 cyanobacterial genera have different responses to nutrients (Figure S5) as previously observed
303 (Anderson *et al* 2015; Harke *et al*., 2016). *M* relative abundance was positively correlated with
304 DP (Figure S5; R^2=7%), in agreement with previous observations (Homma *et al*., 2008). In
305 contrast, *D* was not significantly correlated with DP, and instead was positively correlated with
306 PP and PN (Figure S5). *D* also responded negatively to dissolved nitrogen, in agreement with
307 previous studies demonstrating that Nostocales (the Order containing *Dolichospermum*) are
308 favoured under conditions of low dissolved inorganic nitrogen, due to their ability to utilize
309 atmospheric nitrogen for growth (Suikkanen *et al*., 2013; Andersson *et al*., 2015). Overall, these
310 results confirmed that *M* and *D* share a spatio-temporal niche during a bloom, but have distinct
311 nutrient preferences.

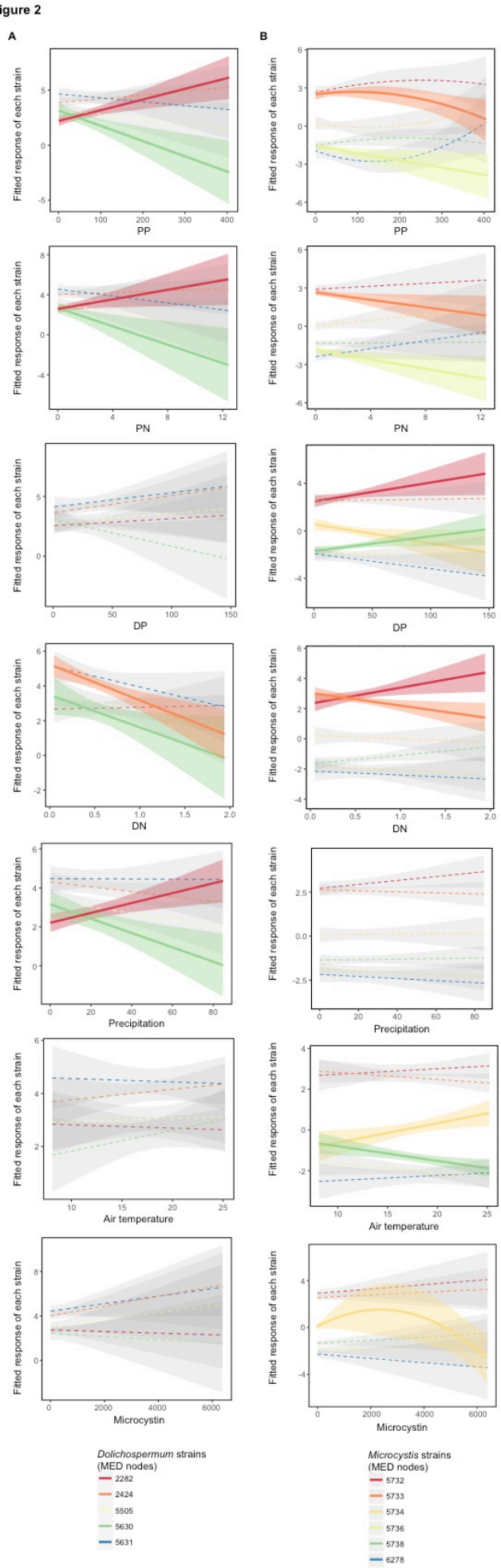
312 **How similar are the niches of strains within a genus?**

313 Numerous studies have focused on environmental conditions that favour cyanobacterial blooms,
314 but few studies have examined how different cyanobacterial strains might respond differently to
315 environmental conditions due to niche partitioning. We observed that *D* and *M* strains, within
316 each genus, appear to have qualitatively different dynamics (Figure S3, S4). To test whether these
317 different dynamics were due to niche partitioning, we first analyzed how the different strains
318 within each genus were related to environmental variables, and then used Latent Variable Models
319 (LVMs) to determine the co-responses (niche similarity) as well as niche separation between
320 strains.

321 We found that nearly all (19/20) of the significant relationships between strain relative
322 abundances (within each genus) and environmental conditions were linear (Figure 2; Table S2).
323 *D* nodes D2282 (red) and D5630 (green) showed opposite responses to nutrients (PP and PN) and
324 precipitation (Figure 2A). The LVMs confirmed a significant niche separation for these niche
325 dimensions (Figure S6). In contrast, nodes D2424 and D5630 both displayed a similar negative
326 relationship (Figure 2A), resulting in a significant co-response (niche similarity) to DN (Figure
327 S6). Within the *M* genus, we observed several nodes with similar responses to PP, PN and DP
328 (Figure 2B), and significant co-responses between nodes were detected for these niches (Figure
329 S7). We also found significant niche separations for DN (involving M5732, M5733 and M5738)
330 and temperature (involving M5734 and M5738) (Figure S7). Overall, the LVM analysis showed
331 that niche separation occurs among nodes within a genus (Figure S8). In *D*, niche separation
332 occurred mainly in the niche dimensions of particulate nutrients (PP and PN) and precipitation
333 (Figure S6), whereas in *M* niche separation occurred mainly according to temperature and DN
334 preferences (Figure S7).

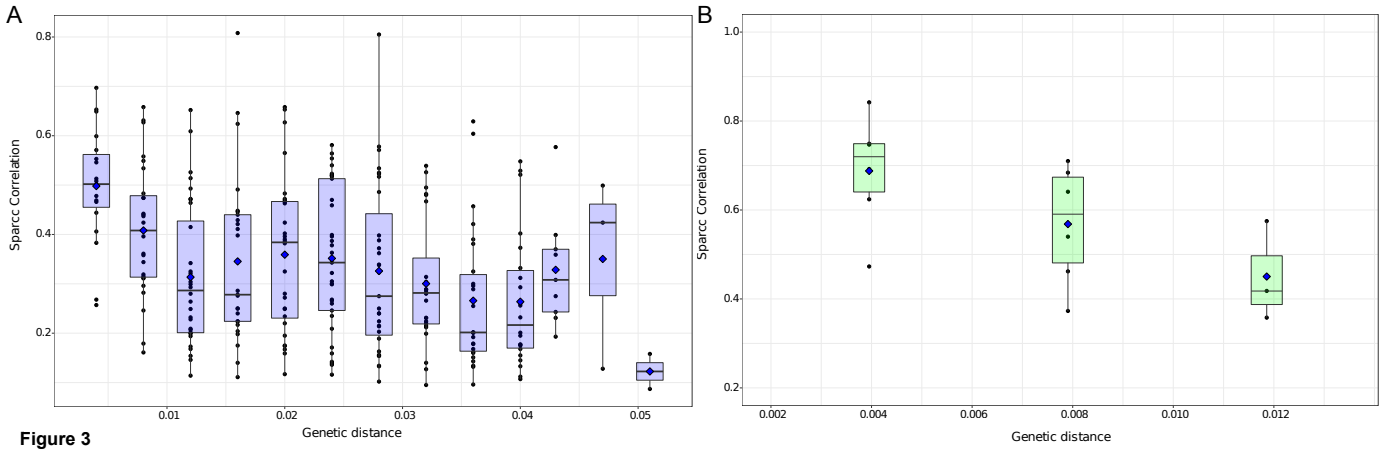
335

Figure 2. Niche partitioning at the sub-genus level. Best-fit polynomial models of the response of *Dolichospermum* (A) and *Microcystis* (B) nodes to abiotic factors. The relative abundance of each MED node (strain) was computed relative to the total number of reads within each genus using the centered-log ratio (clr) transform. Significant relationships are shown by solid lines and coloured confidence intervals. In most cases, the degree-1 polynomial (linear model) provided the best-fit (see Table S1 for details). For *D*, only the dominant nodes (observed in at least 70% of samples) are shown.



362 **How do niche preferences vary with genetic distance?**

363 When two taxa share the same realized niche (*i.e.* where they are actually able to survive in the
364 wild, including biotic and abiotic niches), they are more likely to be observed together, and thus
365 to be correlated in survey data. We therefore used co-occurrence patterns between pairs of *D* or
366 *M* nodes as a proxy for similarity in their realized niches, and asked if more genetically similar
367 nodes are more likely to have similar realized niches. Indeed, we found that pairwise SparCC
368 correlation coefficients between nodes tends to decline with genetic distance (Figure 3).



369

370 **Figure 3. Co-occurrence of strains declines with their pairwise genetic distance.** Relationship
371 between co-occurrence (significant SparCC correlation, $P < 0.05$) and genetic distance (p
372 distance) between *Dolichospermum* (A) and *Microcystis* (B) nodes. Blue diamonds represents the
373 mean SparCC correlation for each distance. Boxplots show the median (horizontal line), the 25th
374 and 75th percentile (enclosed in box) and 95% confidence intervals (whiskers). The discreteness
375 observed in the x-axis is due to the discrete number of substitutions in the 16S rRNA gene
376 sequence (*e.g.* exactly 1, 2, 3,... nucleotide differences between pairs).

377

378 This pattern was significant within both *D* (linear regression, $F(1,279) = 28.3$, $P < 0.001$, adjusted
379 $R^2 = 8.9\%$) and *M* (linear regression, $F(1,13) = 7.9$, $P < 0.05$, adjusted $R^2 = 33.0\%$). The higher
380 R^2 observed for *M* might be explained by its more limited genetic diversity (maximum pairwise

381 genetic distance ~0.01) compared to *D* (maximum distance ~0.05). There appears to be a rapid
382 decline in niche similarity as genetic distance goes from 0 to ~0.01 (adjusted R^2 = 23.0% in *D*
383 when considering only this distance range) followed by a flatter relationship for genetic distance
384 >0.01.

385 Ecologically distinct lineages are expected to be associated with distinct surrounding
386 communities, due to a combination of direct microbe-microbe (biotic) interactions and shared
387 preferences for abiotic conditions (Cohan and Koeppel, 2008). We therefore analyzed the co-
388 occurrence patterns between each *D* or *M* node and other bacterial taxa in the lake community.
389 We identified non-cyanobacterial MED nodes that co-occurred with each *D* and *M* node and
390 found that *M* nodes were generally more connected with other members of the bacterial
391 community (Figure S9). We also found that relatively few taxa (4 out of 26; Table S5) are
392 significantly correlated with both *M* and *D,* suggesting that *M* and *D* strains co-occur with distinct
393 sets of other bacteria, and thus have distinct realized niches.

394 We further investigated whether more closely-related *M* or *D* nodes have more similar
395 correlations with potentially interacting community members. We focused on members of the
396 *Cytophagaceae* family (MED nodes 3667, 5983, and 5984), which are potential predators of
397 cyanobacteria (Rashidan and Bird, 2001), members of the *Rhizobiales* order (nodes 4737, 3705,
398 and 3726), which are potential N-fixers that could provide nitrogen for non-N-fixing *Microcystis*
399 (Louati *et al*., 2015), and all the taxa that co-occurred with both *D* and *M* nodes (nodes 1061,
400 4674, 7272, and 4756). For each pair of *D* or *M* nodes, we correlated their genetic distance (as in
401 Figure 3) with the absolute difference in the correlations (r) with each potentially interacting
402 community member (Figure S11). We found that closely-related *D* nodes tend to have more
403 similar correlations with *Cytophagaceae* node 3667, whereas more distantly-related *D* nodes
404 have more different correlations with node 3667 (Correlation between $|\Delta r|$ and genetic distance,

405 adjusted-R^2=0.059, $P < 0.001$). Similar but not statistically significant patterns were observed for
406 several *M* nodes. It is difficult to generalize from these results, but at least in some cases, more
407 phylogenetically similar strains may share conserved interactions with other community
408 members.
409
410 Finally, we asked whether the decline in niche similarity with genetic distance (Figure 3)
411 was a common feature of all niche dimensions, or if different abiotic parameters showed different
412 patterns. We considered all significant models, both linear and non-linear as shown in Figure 4
413 (for *D*) and Figure S10 (for *M*). We found a negative linear relationship between DN niche
414 similarity and genetic distance (adjusted R^2=25%, $P = 0.0791$). Qualitatively, most other niche
415 dimensions showed a similar pattern for *D* (Figure 4), and among *M* nodes we found a significant
416 negative linear relationship between the correlated fitted response to microcystin concentrations
417 and genetic distance (Figure S10). However, we also identified non-linear relationships between
418 *D* niche similarity and genetic distance for PP (adjusted R^2=69%, $P = 0.0184$) and PN (adjusted
419 R^2=63%, $P = 0.0280$). For these niche dimensions, the correlation of fitted responses declines
420 from genetic distances of 0 to ~0.01, then rises to another peak around ~0.03 before declining
421 again (Figure 4). While these non-linear models provided significantly better fits to the data
422 compared to linear models, we cannot exclude the possibility that the fits were driven by a few
423 outlying points particular to our data set. Replicating these findings in additional data sets (*e.g.*
424 from different lakes, over different time scales) is therefore essential.

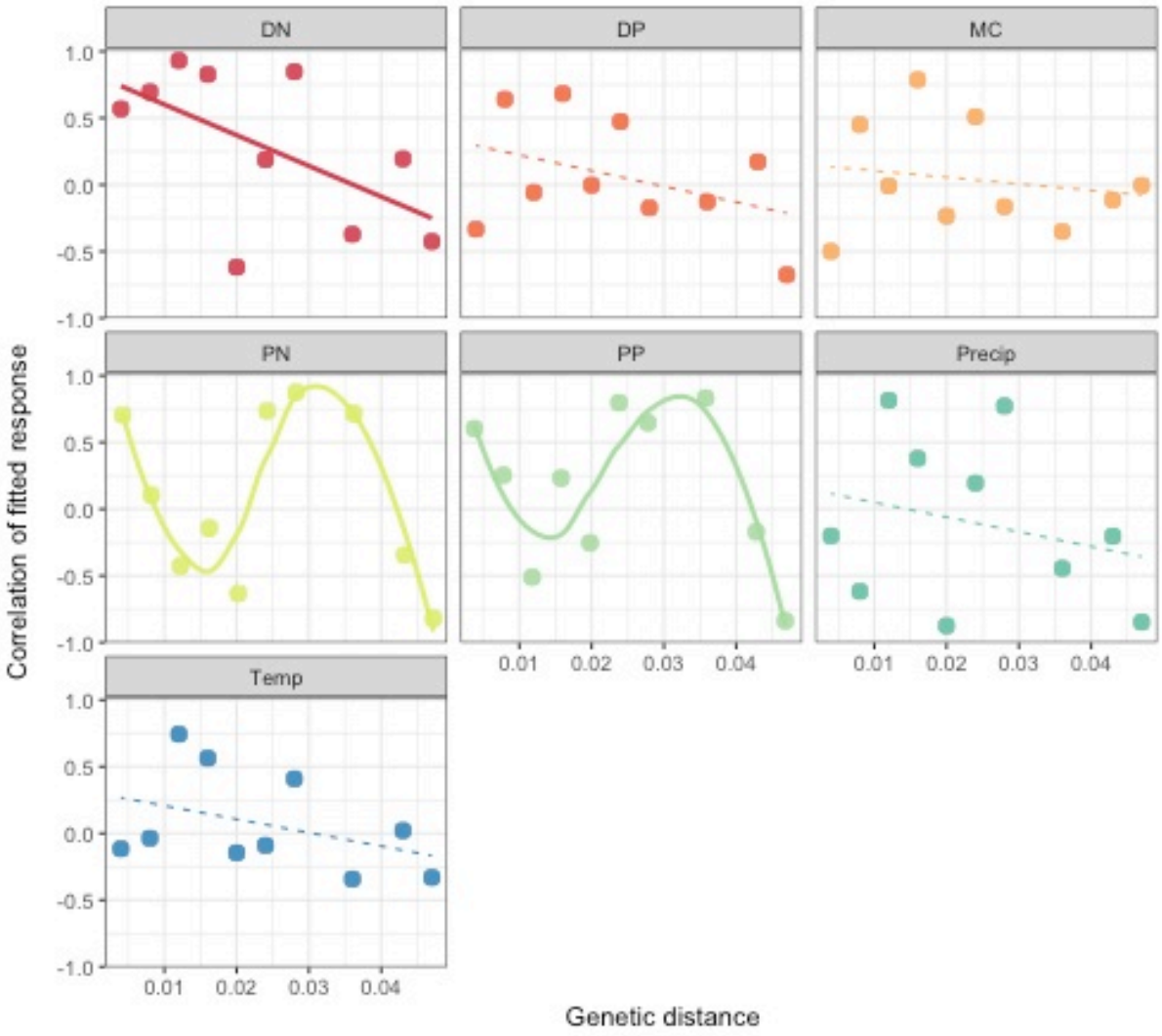


Figure 4. The relationship between genetic distance and the co-response (niche similarity) for *Dolichospermum*. LVMs were used to identify correlations between the responses of MED nodes to each measured environmental parameter (separate panels). Positive correlations of fitted responses indicate similar niches; negative correlations indicate different niche preferences. Genetic distances were computed using the p-distance. Separate model fits were tested with the Akaike information criterion (AIC) for the relationship between each niche dimension and genetic distance. See Table S4 for details of model fits. Significant model fits are shown with thick solid lines; non-significant fits are shown with dashed lines.

436 **Discussion**

437 Overall, our results show how *Microcystis* and *Dolichospermum* can achieve similar levels of
438 dominance during cyanobacterial blooms by partitioning niche space within and between genera.
439 Over eight years of sampling, we observed that *Dolichospermum* and *Microcystis* generally co-
440 occurred during blooms, suggesting a broadly similar physico-temporal niche (Figure 1; Figure
441 S1). However, we confirmed that *D* was associated with lower concentrations of dissolved
442 nitrogen, consistent with its known ability to fix nitrogen (Harke *et al*., 2016, Andersson *et al*.,
443 2015) while *M* was associated with higher concentrations of dissolved phosphorus (Figure S5).
444 These genus-level traits could mask niche differentiation among strains within each genus.

445 To dissect niche preferences at finer taxonomic resolution, below the genus level, we used
446 minimum entropy decomposition (MED), allowing single-nucleotide resolution of 16S amplicon
447 sequences, *i.e.* each MED node is an exact sequence variant. Finer taxonomic resolution has been
448 shown to increase the power to correctly identify phylogenetic overdispersion, a signature of
449 competitive interactions being more important than habitat filtering (Koeppel and Wu, 2014).
450 However, even at high resolution, the 16S marker gene may be too slow-evolving to be a good
451 marker for fast-evolving traits, such as toxin production (Berry *et al*., 2017) – a trait which likely
452 evolves rapidly by horizontal gene transfer (Moffit and Neilan, 2004).

453 Despite these limitations, MED analysis provided additional insights into the distinct
454 niches of *M* and *D*. *Dolichospermum* diversity was higher and most of the MED nodes (strains)
455 were conditionally rare, some of them being bloom-associated in some years but not in others
456 (e.g D2282, Figure S3). *Microcystis* nodes, on the other hand, were less diverse but more
457 consistently bloom-associated (nodes M5732 and M5733; Figure S4) and more correlated with
458 other taxa (Figure S9), suggesting that the two dominant bloom-forming genera might use

459 different ecological strategies, beyond what is already known about nitrogen and phosphorus
460 utilization.

461 Examining niche partitioning below the genus level also allowed us to detect patterns not
462 evident at the genus level. For example, *Dolichospermum* is associated with higher
463 concentrations of PP and PN at the genus level (Figure S5), but contains a strain (D5630) that
464 shows the opposite pattern (Figure 2, Figure S6). Within *Microcystis*, niche partitioning occurred
465 mainly for temperature and DN, and the two most dominant *M* nodes had a significant niche
466 separation for DN (Figure 2; Figure S7). It is known that *Microcystis* does not fix atmospheric
467 nitrogen, but is able to use refractory N-containing compounds such as urea or amino acids
468 (Moisander *et al*., 2009; Dai *et al*., 2009), and DN is likely needed for toxin production
469 (Monchamp *et al*., 2014). Therefore, different *M* strains could specialize in their preference for
470 different forms of nitrogen.

471 The relative importance of habitat filtering and competition in shaping microbial
472 communities is widely debated, and distinguishing between the two processes can be technically
473 challenging (Koeppel and Wu, 2014; Cadotte and Tucker 2017). Consistent with a general effect
474 of habitat filtering in selecting for genetically similar cyanobacteria under similar conditions, we
475 found a negative relationship between MED node co-occurrence and pairwise genetic distance
476 (Figure 3). This result is in agreement with an early study showing the importance of habitat
477 filtering in microbial communities (Horner-Devine and Bohannan, 2006) and is also consistent
478 with what was observed by Silverman *et al* (2017) using human microbiome data. Here we
479 considered co-occurrence as a proxy for shared realized niches, including both biotic and abiotic
480 factors.

481 Consistent with the general importance of habitat filtering in shaping cyanobacterial
482 communities, we observed that closely-related strains tended to have similar co-responses to

483 several measured abiotic environmental parameters. For example, we found negative linear
484 relationships between niche similarity and *Dolichospermum* genetic distance for most niche
485 dimensions, particularly DN (Figure 4). This observation is in agreement with a previous study
486 that showed that bacterial and fungal responses to N fertilization tend to be more similar among
487 close relatives, with a decline in similarity between genetic distances of 0 and 0.05 (Martiny *et*
488 *al*., 2015; Amend *et al*., 2016). In *Microcystis,* we found that more genetically similar strains
489 tended to be observed at more similar concentrations of the toxin microcystin (Figure S10).
490 Although 16S is a poor marker for microcystin production (Berry *et al*., 2017), our results
491 suggest that phylogenetic relatedness is nevertheless somewhat predictive of microcystin
492 concentrations. This means that microcystin production or tolerance is more likely to be shared
493 by close relatives.

494 In contrast to the general pattern of declining niche similarity with genetic distance, PN
495 and PP both had non-linear relationships with genetic distance in *Dolichospermum* (Figure 4).
496 After an initial decline in co-responses to PN and PP to a genetic distance of ~0.01, co-response
497 rose to another peak at ~0.03 before declining again. This response is consistent with near-
498 identical MED nodes sharing the same preferences for PN and PP concentrations, and that
499 competition between close-relatives (up to genetic distance of ~0.01) imposes divergent selection
500 for distinct niche preferences. The similarity in PN and PP niches for more distant relatives
501 (distance ~0.03) can be explained if distant relatives have diverged to reduce competition in other
502 niche dimensions, allowing them to converge in PN and PP preferences. It is unclear why this
503 non-linear relationship is observed for PN and PP, but not for other niche dimensions. One
504 possibility is that P uptake and metabolism genes are easily acquired by horizontal gene transfer,
505 as observed in marine cyanobacteria (Coleman *et al*., 2010) and may thus be more rapidly
506 evolving, resulting in non-linear relationship with phylogenetic distance. Yet it is unclear why the

507 non-linear pattern is observed for particulate but not dissolved N and P. Particulate nutrients
508 could be a marker for biomass (*e.g.* bloom density), but it is equally unclear how rapidly "bloom
509 preference" would be expected to evolve. The same non-linear pattern might be expected for
510 microcystin, due to frequent gain/loss of the underlying biosynthetic genes. However, we observe
511 a linear relationship (Figure S10) between microcystin concentrations and genetic distance
512 among *Microcystis* strains – but only in a very narrow range of genetic distance (99-100%
513 identity). Therefore, non-linear relationships could become apparent among more distantly
514 related strains (*e.g.* 95-99% identity).

515 To investigate how biotic niches change over evolutionary time, we studied how
516 cyanobacterial interactions (SparCC correlations) with non-cyanobacterial community members
517 varied with genetic distance among cyanobacteria. We found that genetically similar
518 *Dolichospermum* strains tended to have similar interactions with a strain of *Cytophagaceae*, a
519 potential cyanobacterial predator, whereas more distantly related *Dolichospermum* strains
520 diverged in terms of their interactions (Figure S11). We limited our analyses to only the most
521 common interacting partners (common to *M* and *D*) or those previously suspected to interact with
522 cyanobacteria via predation or cross-feeding. Future studies could expand on these analyses in a
523 more comprehensive fashion to quantify how biotic interactions evolve over different time scales
524 and in different lineages.

525 Overall, our results show how different traits may have different relationships with
526 genetic distance, highlighting the importance of considering each niche dimension separately,
527 because adaptation to different niche dimensions can occur at dramatically different rates
528 (Martiny *et al*., 2015). Our results also suggest that the same trait could evolve at different rates
529 in different cyanobacterial lineages. For example, temperature adaptation evolves at an
530 intermediate rate in *Prochloroccocus* (Martiny *et al.*, 2015) and perhaps at a slower rate in

531 hotspring *Synechococcus* (Becraft *et al*., 2011; Becraft *et al*., 2015). Using our comparative
532 framework of two broadly sympatric genera, we identify a general decline in ecological similarity
533 with genetic distance – although certain traits (*e.g.* PN and PP in *Dolichospermum* but not in
534 *Microcystis*) go against this trend. A future challenge in microbial ecology and evolution will be
535 to determine which traits are generally fast- or slow-evolving, and most interestingly, which
536 lineages provide exceptions. For example, particularly rapid evolution of a generally slow-
537 evolving trait in a particular lineage could provide evidence for an unusual genetic architecture or
538 strong selective pressure on that particular trait in that particular lineage.

References

Aitchison, J. (1986) The statistical analysis of compositional data. Chapman and Hall, London, 416 p. doi.org/10.1007/978-94-009-4109-0

Amend, A. S., Martiny, A. C., Allison, S. D., Berlemont, R., Goulden, M. L., Lu, Y., *et al*. (2016). Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. *ISME J* 10, 109–118. doi:10.1038/ismej.2015.96.

Andersson, A., Höglander, H., Karlsson, C., and Huseby, S. (2015). Key role of phosphorus and nitrogen in regulating cyanobacterial community composition in the northern Baltic Sea. *Estuar Coast Shelf Sci.* 164, 161–171. doi:10.1016/j.ecss.2015.07.013.

Becraft E. D., Cohan F. M., Kuhl M., Jensen S. I., Ward D. M. (2011). Fine-scale distribution patterns of Synechococcus ecological diversity in microbial mats of Mushroom Spring, Yellowstone National Park. Appl. Environ. Microbiol. 77 7689–7697. 10.1128/AEM.05927-5911

Becraft, E. D., Wood, J. M., Rusch, D. B., Kühl, M., Jensen, S. I., Bryant, D. A., et al. (2015). The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of Synechococcus inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. Front Microbiol 6. doi:10.3389/fmicb.2015.00590.

Berry, M. A., White, J. D., Davis, T. W., Jain, S., Johengen, T. H., Dick, G. J., *et al*. (2017). Are oligotypes meaningful ecological and phylogenetic units? A case study of *Microcystis* in freshwater lakes. *Front Microbiol* 8. doi:10.3389/fmicb.2017.00365.

Cadotte, M. W., and Tucker, C. M. (2017). Should environmental filtering be abandoned? Trends Ecol. Evol. 32, 429–437. doi:10.1016/j.tree.2017.03.004.

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., and Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715. doi:10.1111/j.1461-0248.2009.01314.x.

Cohan, F. M., and Koeppel, A. F. (2008). The origins of ecological diversity in Prokaryotes. *Current Biology* 18, R1024–R1034. doi:10.1016/j.cub.2008.09.014.

Coleman, M. L., and Chisholm, S. W. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18634–18639. doi:10.1073/pnas.1009480107.

Dai, R., Liu, H., Qu, J., Zhao, X., and Hou, Y. (2009). Effects of amino acids on microcystin production of the *Microcystis aeruginosa*. *J. Hazard. Mater.* 161, 730–736. doi:10.1016/j.jhazmat.2008.04.015.

Dolman, A. M., Rücker, J., Pick, F. R., Fastner, J., Rohrlack, T., Mischke, U., *et al*. (2012). Cyanobacteria and Cyanotoxins: The influence of Nitrogen versus Phosphorus. *PLoS One*

7, e38757. doi:10.1371/journal.pone.0038757.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi:10.1093/bioinformatics/btq461.

Eren, A. M., Morrison, H. G., Vineis, J. H., Reveillaud, J., Sogin, M. L., and Lescault, P. J. (2014). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* 9, 968. doi:10.1038/ismej.2014.195.

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Computational Biology* 8, e1002687. doi:10.1371/journal.pcbi.1002687.

Gauze, G. F. (1934). *The struggle for existence*. Baltimore, The Williams & Wilkins company Available at: http://archive.org/details/struggleforexist00gauz

Gloor, G. B., and Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can. *J. Microbiol*. 62, 692–703. doi:10.1139/cjm-2015-0821.

Gray, N. D., Comaskey, D., Miskin, I. P., Pickup, R. W., Suzuki, K., and Head, I. M. (2004). Adaptation of sympatric *Achromatium* spp. to different redox conditions as a mechanism for coexistence of functionally similar sulphur bacteria. *Environ. Microbiol.* 6, 669–677. doi:10.1111/j.1462-2920.2004.00607.x.

Hajdu, S., Höglander, H., and Larsson, U. (2007). Phytoplankton vertical distributions and composition in Baltic Sea cyanobacterial blooms. *Harmful Algae* 6, 189–205. doi:10.1016/j.hal.2006.07.006.

Harke, M. J., Davis, T. W., Watson, S. B., and Gobler, C. J. (2016). Nutrient-controlled niche differentiation of western lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environ. Sci. Technol.* 50, 604–615. doi:10.1021/acs.est.5b03931.

Homma, T., Komatsu, N., Negishi, M., Katagami, Y., Nakamura, K., and Park, H. D. (2008). Influence of dissolved inorganic nitrogen and phosphorus concentrations on the horizontal and temporal changes of *Microcystis* population in lake Kitaura. The XII World Lake Conference, Taal. 1423-1429.

Horner-Devine, M. C., and Bohannan, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* 87, S100-108.

Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D., and Warton, D. I. (2015). Model-based approaches to unconstrained ordination. *Methods Ecol Evol* 6, 399–411. doi:10.1111/2041-210X.12236.

Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., and Polz, M. F. (2008).

Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320, 1081–1085. doi:10.1126/science.1157890.

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harb Symp Quant Biol* 22, 415–427. doi:10.1101/SQB.1957.022.01.039.

Jaspers, E., and Overmann, J. (2004). Ecological significance of microdiversity: identical 16s rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl Environ Microbiol* 70, 4831–4839. doi:10.1128/AEM.70.8.4831-4839.2004.

Jezbera, J., Jezberová, J., Brandt, U., and Hahn, M. W. (2011). Ubiquity of *Polynucleobacter* necessarius subspecies asymbioticus results from ecological diversification. *Environ Microbiol* 13, 922–931. doi:10.1111/j.1462-2920.2010.02396.x.

Koeppel, A. F., and Wu, M. (2014). Species matter: the role of competition in the assembly of congeneric bacteria. *ISME J* 8, 531–540. doi:10.1038/ismej.2013.180.

Konopka, A., and Brock, T. D. (1978). Effect of temperature on blue-green algae (Cyanobacteria) in Lake Mendota. *Appl Environ Microbiol* 36, 572–576.

Letten, A. D., Keith, D. A., Tozer, M. G., and Hui, F. K. C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *J Ecol* 103, 1264–1275. doi:10.1111/1365-2745.12428.

Louati, I., Pascault, N., Debroas, D., Bernard, C., Humbert, J.-F., and Leloup, J. (2015). Structural diversity of bacterial communities associated with bloom-forming freshwater cyanobacteria differs according to the cyanobacterial Genus. *PLoS One* 10, e0140614. doi:10.1371/journal.pone.0140614.

Martiny, J. B. H., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science* 350, aac9323. doi:10.1126/science.aac9323.

Moffit, M.C. and Neilan, B.A. (2004). Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl. and Environ, Microbiol.* 70, 6353-6362

Moisander, P. H., Ochiai, M., and Lincoff, A. (2009). Nutrient limitation of *Microcystis aeruginosa* in northern California Klamath River reservoirs. *Harmful Algae* 8, 889–897. doi:10.1016/j.hal.2009.04.005.

Monchamp, M.-E., Pick, F. R., Beisner, B. E., and Maranger, R. (2014). Nitrogen forms influence microcystin concentration and composition via changes in cyanobacterial community structure. *PLoS One* 9, e85573. doi:10.1371/journal.pone.0085573.

Neuenschwander, S. M., Pernthaler, J., Posch, T., and Salcher, M. M. (2015). Seasonal growth potential of rare lake water bacteria suggest their disproportional contribution to carbon fluxes. *Environ Microbiol* 17, 781–795. doi:10.1111/1462-2920.12520.

647 Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., and Bertilsson, S. (2011). A guide to the
648 natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* 75, 14–49.
649 doi:10.1128/MMBR.00028-10.

650 Olli, K., Klais, R., and Tamminen, T. (2015). Rehabilitating the cyanobacteria – niche
651 partitioning, resource use efficiency and phytoplankton community structure during
652 diazotrophic cyanobacterial blooms. J Ecol 103, 1153–1164. doi:10.1111/1365-
653 2745.12437.

654 Paerl, H. W., Fulton, R. S., Moisander, P. H., and Dyble, J. (2001). Harmful freshwater algal
655 blooms, with an emphasis on cyanobacteria. *Scientific World Journal* 1, 76–113.
656 doi:10.1100/tsw.2001.16.

657 Paliy O. and Shankar ,V. (2016). Application of multivariate statistical techniques in microbial
658 ecology. *Mol Ecol* 25,1032–1057. doi:10.1111/mec.13536.

659 Rantala, A., Fewer, D. P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Börner, T., *et al*. (2004).
660 Phylogenetic evidence for the early evolution of microcystin synthesis. Proc. Natl. Acad.
661 Sci. U.S.A. 101, 568–573. doi:10.1073/pnas.0304489101.

662 Rashidan, K. K., and Bird, D. F. (2001). Role of predatory bacteria in the termination of a
663 cyanobacterial bloom. *Microb. Ecol.* 41, 97–105. doi:10.1007/s002480000074.

664 Rohwer, R. R., Hamilton, J. J., Newton, R. J., and McMahon, K. D. (2017). Taxass: Leveraging
665 custom databases achieves fine-scale taxonomic resolution. *bioRxiv*, 214288.
666 doi:10.1101/214288.

667 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., *et al*.
668 (2009). Introducing Mothur: open-source, platform-independent, community-supported
669 software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*
670 75, 7537–7541. doi:10.1128/AEM.01541-09.

671 Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., *et al*. (2014).
672 Conditionally rare taxa disproportionately contribute to temporal changes in microbial
673 diversity. *mBio* 5, e01371-14. doi:10.1128/mBio.01371-14.

674 Tromas, N., Fortin, N., Bedrani, L., Terrat, Y., Cardoso, P., Bird, D., *et al*. (2017). Characterizing
675 and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course.
676 *ISME J* 11, 1746–1763. doi:10.1038/ismej.2017.58

677 Silverman, J. D., Washburne, A. D., Mukherjee, S., and David, L. A. (2017). A phylogenetic
678 transform enhances analysis of compositional microbiota data. *eLife* 6, e21887.
679 doi:10.7554/eLife.21887.

680 Šimek, K., Kasalický, V., Horňák, K., Hahn, M. W., and Weinbauer, M. G. (2010). Assessing
681 niche separation among coexisting *limnohabitans* strains through interactions with a
682 competitor, viruses, and a bacterivore. *Appl. Environ. Microbiol.* 76, 3762–3762.
683 doi:10.1128/AEM.00774-10.

Suikkanen, S., Pulina, S., Engström-Öst, J., Lehtiniemi, M., Lehtinen, S., and Brutemark, A. (2013). Climate change and eutrophication induced shifts in northern summer plankton communities. *PLoS One* 8, e66475. doi:10.1371/journal.pone.0066475.

Tilman, D. (1982). *Resource Competition and Community Structure*. Princeton University Press.

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., *et al*. (2015). So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution* 30, 766–779. doi:10.1016/j.tree.2015.09.007.

Willis, A., and Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics* 71, 1042–1049. doi:10.1111/biom.12332.

Willis, A., Bunge, J., and Whitman, T. (2017). Improved detection of changes in species richness in high diversity microbial communities. J. R. Stat. Soc. C 66, 963–977. doi:10.1111/rssc.12206.

Yamamoto, Y., and Tsukada, H. (2009). Measurement of in situ specific growth rates of *Microcystis* (cyanobacteria) from the frequency of dividing cells. *J. Phycol.* 45, 1003–1009. doi:10.1111/j.1529-8817.2009.00723.x.

Data availability

Raw sequence data have been deposited NCBI GenBank under BioProject number PRJNA353865: SRP099259.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We thank Francis Hui, Andrew Letten and David Warton for advice on the LVM analyses, and the two peer reviewers for their constructive comments that improved the manuscript. We also thank everyone who participated in sampling, data collection and analysis. This research was funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery grant and a Fonds de Recherche du Québec Nature et Technologies (FRQNT) New Researcher grant to BJS, a FRQNT Programme de recherche en partenariat sur les cyanobactéries grant to CWG, and the federal government interdepartmental Genomics Research and Development Initiative (GRDI).

NT is funded by a project from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 656647.