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 cyanobacteria 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). 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.

INTRODUCTION

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

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

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

74 In this study, we use Hutchinson's definition of a fundamental niche as the set of abiotic
75 conditions under which an organism can survive and reproduce, and a realized niche as the
76 conditions under which it is actually observed in practice, accounting for both abiotic and biotic
77 (*e.g.* competition, predation, cooperation) interactions (Hutchinson 1957). If two species have
78 identical ecological niches, one should competitively exclude the other (Gause 1934; Tilman
79 1982) unless competition is weak due to abundant resources. In practice, closely related taxa
80 often compete for space and resources (Cavender-Bares et al., 2009), favouring specialization to
81 reduce overlap in niche space. For example, coexisting (sympatric) taxa within the same genus
82 tend to have different realized niches, experiencing different seasonal growth patterns or
83 responding differently to environmental parameters (Gray et al., 2004; Jaspers and Overmann,
84 2004; Hunt et al, 2008; Simek et al, 2010; Jezberra et al, 2011; Neuenschwander et al., 2015).
85 Pairs of taxa with similar realized niches can be identified when they co-occur in repeated
86 sampling over space and time, and such co-occurrence networks are readily inferred from deep
87 amplicon sequencing datasets (Friedman and Alm, 2012).

88 Cyanobacteria are widely and naturally present in freshwater ecosystems, and some
89 lineages form blooms under appropriate conditions of temperature and nutrients (Konopka and
90 Broke, 1978; Harke et al, 2015). Several studies have shown that, at the genus level, some bloom-
91 forming cyanobacteria share similar niches (Pearl et al, 2001; Yamamoto, 2009) allowing them to
92 co-occur during bloom events. However, the degree of similarity among ecological niches of co-
93 occurring cyanobacteria (e.g. *Microcystis* and *Dolichospermum*) remains unclear.

94 We investigated niche partitioning among *Microcystis* (*M*) and *Dolichospermum* (*D*) that
95 induce blooms nearly every year in a large eutrophic North American lake, Lake Champlain. In a

96 previously described 8-year time-course analysis (2006 to 2013) spanning multiple bloom events,
97 we used 16S amplicon sequencing to broadly survey changes in the lake microbial community
98 over time, generally at the genus level (Tromas et al., 2017). Here, we use Minimum Entropy
99 Decomposition of amplicon sequences (Eren *et al*., 2015) to identify sub-genus strains (MED
100 nodes; here used interchangeably with "strains") within each of the two dominant cyanobacterial
101 genera, *M* and *D*, at single nucleotide resolution. A previous study demonstrated that oligotypes
102 (similar to MED nodes) lacked the resolution to distinguish toxic and non-toxic *Microcystis*
103 lineages, but could potentially be informative about other, more phylogenetically conserved
104 niches (Berry et al., 2017). We used a combination of genetic data (diversity of MED nodes) and
105 matched environmental data (*e.g.* temperature, nutrient concentrations, precipitation) to address
106 three specific questions. First, how similar are the niches of the two dominant cyanobacterial
107 genera *M* and *D*? Second, how similar are the niches of strains within each genus? Third, how
108 does niche similarity change with genetic relatedness? We confirm that *M* and *D* are broadly co-
109 occuring during blooms, but have distinct nutrient preferences. We also identified niche
110 partitioning at the sub-genus level, and observe a general decline in realized niche similarity with
111 genetic distance, consistent with habitat filtering. However, certain niche dimensions (particulate
112 nutrient concentrations) show a complex polynomial relationship with genetic distance,
113 suggesting that a combination of habitat filtering and competitive interactions shapes the
114 evolution of these traits.
115
116

117 **MATERIALS AND METHODS**

118

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

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

124

125 **Sequence analysis**

126 A total of 7,949,602 sequences were obtained from 150 lake samples, with a median of
127 41,982 per sample as previously described (Tromas et al., 2017). These sequences were
128 processed with the default parameters of the SmileTrain pipeline
129 (https://github.com/almlab/SmileTrain/wiki) that includes reads quality filtering, chimera
130 filtering, and merging using USEARCH (version 7.0.1090, http://www.drive5.com/usearch/,
131 default parameter) (Edgar, 2010), Mothur (version 1.33.3) (Schloss *et al*., 2009), Biopython
132 (version 2.7) and custom scripts. Minimum Entropy Decomposition (MED) was then applied to
133 the filtered and merged reads to partition sequence reads into MED nodes (Eren *et al*., 2015).
134 MED was performed using the following parameters –M noise filter set to 500 resulting in ~7%
135 of reads filtered and 942 MED nodes. Samples with less than 1000 reads were removed, yielding
136 a final dataset of 135 samples. Finally, taxonomy was assigned using the assign_taxonomy.py
137 QIIME script (default parameters), and a combination of GreenGenes and a freshwater-specific
138 database (Freshwater database 2016 August 18 release; Newton *et al*., 2011), using TaxAss
139 (https://github.com/McMahonLab/TaxAss, installation date: September 13th 2016; Rohwer et al.,
140 2017). After assignment, nodes that belong to Eukaryotes but still present in the database

141 (Cryptophyta, Streptophyta, Chlorophyta and Stramenopiles orders) were removed, leading to a
142 total of 891 nodes.
143
144 **Diversity analysis**
145 Comparing changes in the diversity of *M* strains to the diversity of *D* strains was
146 performed using betta (Willis et al., 2017). betta accounts for strains that are present in the
147 environment but not observed in the samples due to incomplete sampling. The number of
148 unobserved strains is estimated based on the number of strains that are observed and their
149 abundances. The total strain diversity was estimated using breakaway (R package v4) which
150 accounts for ecological interactions between strains (Willis and Bunge, 2015).
151
152 **Conditionally rare taxa analysis**
153 We investigated the temporal dynamics of *M* and *D* nodes by measuring the composition for each
154 genus in conditionally rare taxa. The matrix of node absolute abundances was used as input for
155 the R script CRT_Functions_v1.1.R (Shade et al., 2014;
156 https://github.com/ShadeLab/ConditionallyRareTaxa) using the default parameters. Conditionally
157 rare taxa are defined as rare taxa that occasionally could become very abundant, without showing
158 rhythmic or seasonal patterns.
159
160 **Node–environment relationships analysis**
161 To investigate node–environment relationships, we used an environmental data matrix that
162 included: particulate phosphorus in µg/ L (PP, the difference between TP and DP), particulate
163 nitrogen in mg/ L (PN, the difference between TN and DN), soluble reactive phosphorus in µg/ L
164 (DP), dissolved nitrogen in mg/ L (DN), 1-week-cumulative precipitation in mm and 1-week-

165 average air temperature in Celsius. The measurements of the different environment variables are
166 described in Tromas et al (2017). In this previous study, we showed that these environmental
167 variable, over the 8 years, were not correlated with one another.
168
169 **Response to abiotic factors**
170 To analyze the response of each node to abiotic environmental data, we used a Latent
171 Variable Model (LVM) framework, which combines Generalized Linear Models with Bayesian
172 Markov Chain Monte Carlo (MCMC) methods (boral package in R; Hui 2015, Warton *et al.*
173 2015). LVM is a model-based approach for analyzing multivariate data (*e.g.* numerous taxa
174 within the response matrix) that partitions the different drivers of taxa co-occurrence patterns into
175 two components: the first is a regression component, which models the taxa-specific
176 environmental responses, and the second is a latent variable component, which is used to identify
177 residual patterns of co-occurrence resulting from unmeasured factors and/or biotic interactions
178 (Letten *et al.* 2015). In this study, we used LVMs to examine how taxa co-respond to abiotic
179 gradients, and used these co-responses as a proxy for niche similarity. To do so, we extracted the
180 *environmental correlation matrix* from the regression model, which, for any two taxa,
181 corresponds to the correlation between their fitted values ($x_i\beta_j$). In particular, for each node we
182 calculated the predicted probability of mean abundance for each site on linear and nonlinear (*i.e.*
183 $Y \sim X + X^2$) response scales, providing a vector of fitted values. Correlations among the fitted
184 responses of any two taxa were then calculated based upon these vectors (correlating the vector
185 of taxon A with the vector of taxon B, for instance).
186 We first tested LVMs using a response matrix of two columns, corresponding to the
187 relative abundances of the *D* and *M* genera, relative to the rest of the bacterial community. For
188 these models, *D* and *M* abundances were normalized by the total counts of all bacteria using the

189 centered log ratio to correct for data compositionality (clr-inter genera) (Aitchison, 1986; Paliy
190 and Shankar, 2016) using a a zero-replacement procedure as suggested in Gloor and Reid (2016).
191 To then examine dynamics within these genera, we ran separate LVMs for *D* and *M* nodes (*i.e.*
192 response matrix consisted of the *D* and *M* nodes, respectively). For this second set of LVMs,
193 node abundances were normalized by the total counts of *D* and *M*, respectively, to obtain an
194 intra-genera relative abundance (clr-intra genera) for each node. Environmental variables were
195 standardized to mean zero and unit variance to reduce the correlation between the linear (X) and
196 nonlinear (X^2) fit, which also simplified model interpretation and stabilized MCMC sampling. As
197 most of the *D* nodes were conditionally rare, we included only the *D* nodes that were present in at
198 least 70% of samples, to avoid overfitting based on too little data. To complement this
199 multivariate LVM analysis and help visualize the univariate response of each genera and node to
200 the suite of environmental variables, we also tested linear (degree-1 polynomial) and nonlinear
201 (degree-2 polynomial) relationships between each response and explanatory variable. We used
202 AIC to find the single best-fit model (either linear or nonlinear) for each relationship and plotted
203 the best-fit model.

204 Lastly, to test whether co-responses or niche separations were stronger for more closely or
205 distantly related nodes, we extracted the correlations between fitted responses of any two nodes
206 from the *environmental correlation matrix* of each best-fit LVM and examined the relationship
207 between each significant correlation and the pairwise genetic distances of the two respective taxa
208 (percent identity in the V4 region of the 16S gene). Here, significant positive correlations
209 represent co-response between any two taxa whereas significant negative correlations represent
210 the degree of niche separation. The genetic distance between nodes was measured within each
211 genus *M* and *D* using the software MEGA (version 7.0.18) using the p-distance (the proportion of

212 nucleotide sites at which two sequences differ), calculated by dividing the number of sites with
213 nucleotide differences by the total number of sites compared (excluding sites with gaps).
214
215 **Co-occurrence analysis**
216 Co-occurrences between *Microcystis*/*Dolichospermum* and other taxa were calculated
217 with SparCC (Friedman and Alm, 2012), with 20 iterations to estimate the median correlation of
218 each pair of MED nodes, and 500 bootstraps to assess the statistical significance. Correlations
219 were then filtered for statistical significance ($P < 0.01$) and correlations with $R > \pm 0.6$, were
220 selected to build networks using Cytoscape (version 3.1.0).
221
222 **RESULTS**
223
224 **Do *Microcystis* and *Dolichospermum* co-occur temporally?**
225 We have previously shown that *M* and *D* were the two most dominant bloom forming
226 cyanobacteria in Lake Champlain's Missisquoi Bay (Tromas et al., 2017). Both genera were
227 present every year at both littoral and pelagic sampling sites, between 2006 and 2013, with a
228 relative abundance of at least 15% on average during summer (max=24.8%), with the exception
229 of summer 2007 (0.7%) when no substantial blooms occurred (Figure 1). *D* was generally the
230 most dominant genus except for 2009 and 2010 when *M* dominated (Figure S1). *D* and *M* were
231 both present every year, (Figure S1) suggesting that they share the same physico-temporal niche
232 during a bloom, although other aspects of their niches likely differ. For example, *D* can fix
233 nitrogen but *M* cannot (Hajdu et al., 2007).

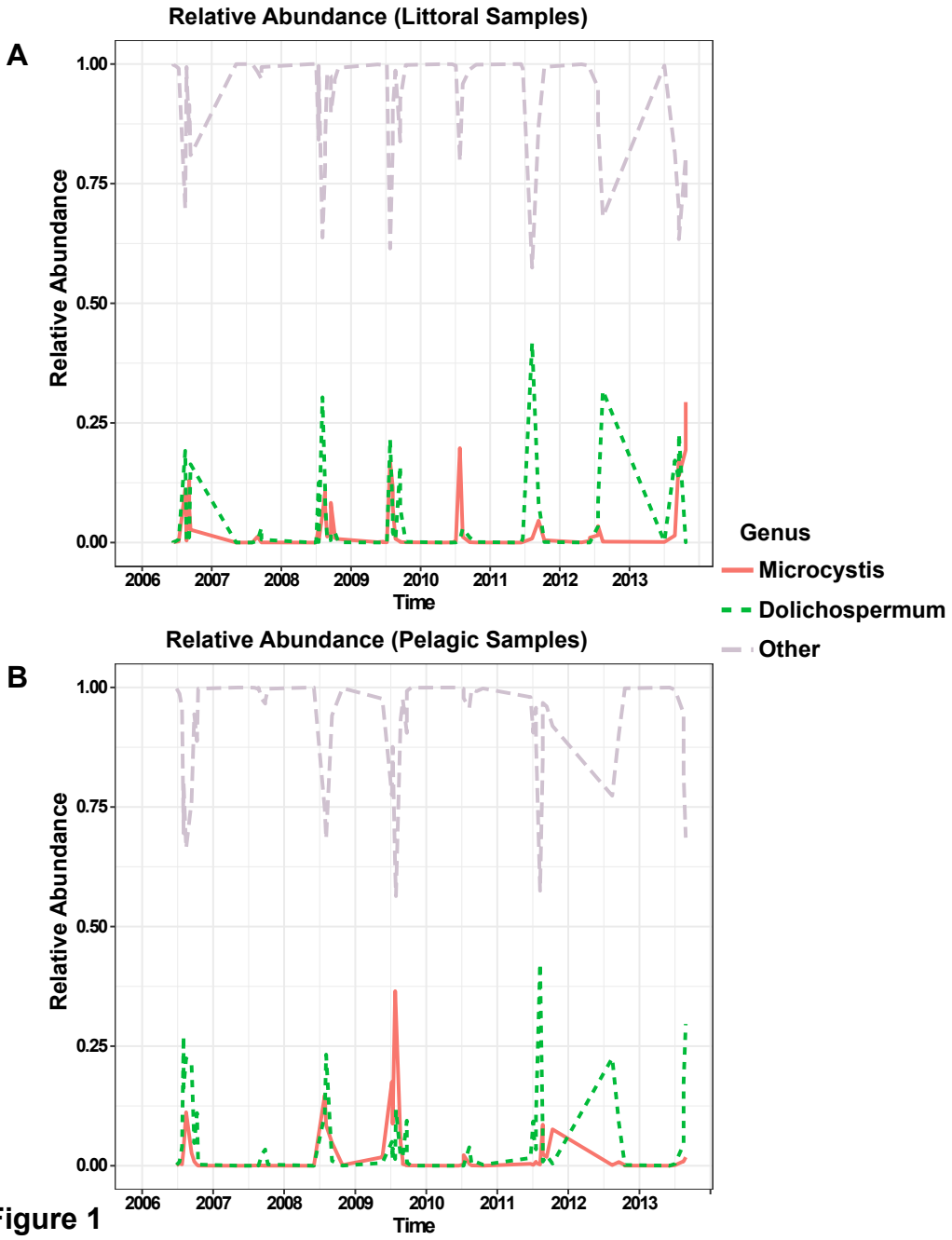


Figure 1. Temporal dynamics of the two dominant cyanobacterial genera over an eight-year time course in littoral (A) and pelagic (B) sampling sites. Relative abundance of *Microcystis* is shown in red, *Dolichospermum* in green and the other members of the bacterial community in blue. The time scale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

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

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

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

253

	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	$P > 0.1$

254

255 **Table 1.** Conditionally rarity analysis. We characterized here the number of MED nodes for two
256 cyanobacterial genera, *Microcystis* and *Dolichospermum* and their dynamic over the 8-years time
257 course by using a conditionally rare analysis (Methods).

258

259 The proportion of rare *D* nodes was significantly higher than what is observed in other genera in
260 the lake community (Fisher test, $P < 0.001$). Several conditionally rare *D* nodes seemed to
261 dominate several times without showing seasonal patterns (Figure S3). Furthermore, we noticed a
262 shift in node composition after 2011, where node D5505 decreased while D2282 increased in
263 relative abundance. In contrast, we only found one conditionally rare *M* node. However, only two
264 *M* nodes (M5732 and M5733) were consistently dominant over time (Figure S4). Similarly, only
265 two *D* nodes, but not always the same two, dominated at any given time (Figure S3). Overall,
266 these results suggest that *D* is more genetically diverse, and that this genetic diversity varies over
267 time. In contrast, *M* is less diverse and more stable over time.
268
269 **Do *Microcystis* and *Dolichospermum* share the same realized niche within the community?**
270 To investigate the niche separation between *D* and *M*, we analyzed their relationships with
271 several environmental conditions measured at the time of sampling (temperature, nutrient
272 concentrations, precipitation and toxin concentrations). We observed that these two
273 cyanobacterial genera have different responses to nutrients (Figure S5) as previously observed
274 (Anderson et al 2015; Harke et al., 2016). *M* relative abundance was positively correlated with
275 DP (Figure S5; R^2=7%), in agreement with previous observations (Homma et al., 2008). In
276 contrast, *D* was not significantly correlated with DP, and instead was positively correlated with
277 PP and PN (Figure S5). *D* also responded negatively to dissolved nitrogen, in agreement with
278 previous studies demonstrating that Nostocales (the Order containing *Dolichospermum*) are
279 favoured under conditions of low dissolved inorganic nitrogen, due to their ability to utilize
280 atmospheric nitrogen for growth (Suikkanen et al., 2013; Andersson et al., 2015). Overall, these
281 results confirmed that *M* and *D* share a spatio-temporal niche during a bloom, but have distinct
282 nutrient preferences.

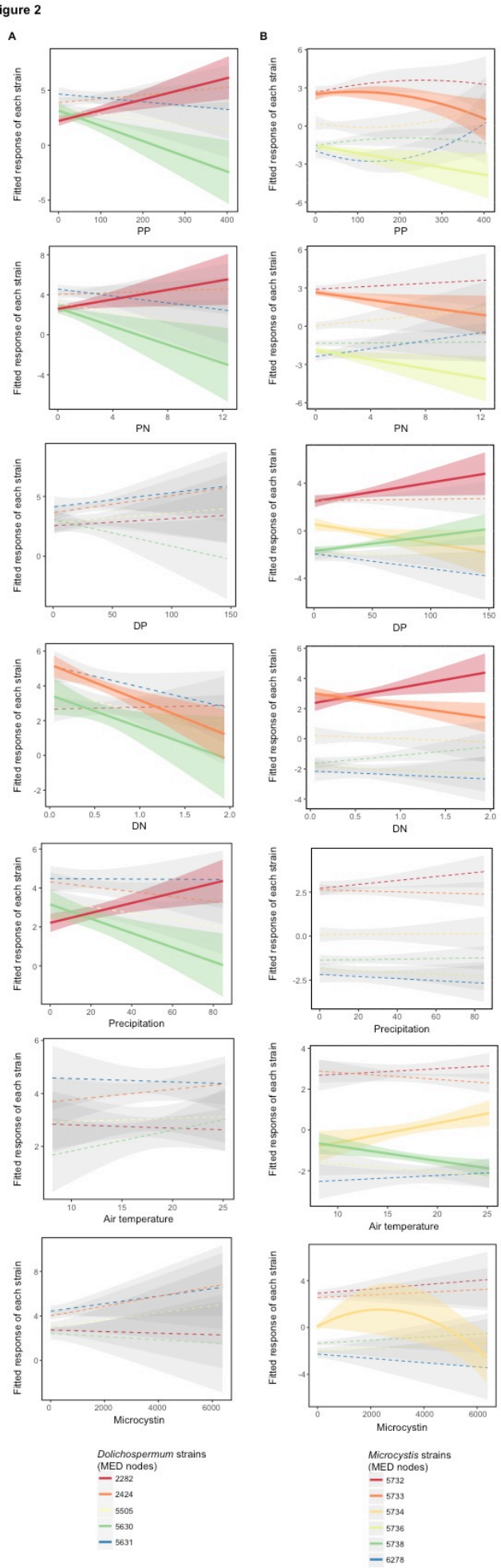
283 **How similar are the niches of strains within a genus?**

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

292 We found that nearly all (19/20) of the significant relationships between strain relative
293 abundances (within each genus) and environmental conditions were linear (Figure 2; Table S2).
294 *D* nodes D2282 (red) and D5630 (green) showed opposite responses to nutrients (PP and PN) and
295 precipitation (Figure 2A). The LVMs confirmed a significant niche separation for these niche
296 dimensions (Figure S6). In contrast, nodes D2424 and D5630 both displayed a similar negative
297 relationship (Figure 2A), resulting in a significant co-response (niche similarity) to DN (Figure
298 S6). Within the *M* genus, we observed several nodes with similar responses to PP, PN and DP
299 (Figure 2B), and significant co-responses between nodes were detected for these niches (Figure
300 S7). We also found significant niche separations for DN (involving M5732, M5733 and M5738)
301 and temperature (involving M5734 and M5738) (Figure S7). Overall, the LVM analysis showed
302 that niche separation occurs among nodes within a genus. In *D*, niche separation occurred mainly
303 in the niche dimensions of particulate nutrients (PP and PN) and precipitation (Figure S6),
304 whereas in *M* niche separation occurred mainly according to temperature and DN preferences
305 (Figure S7).

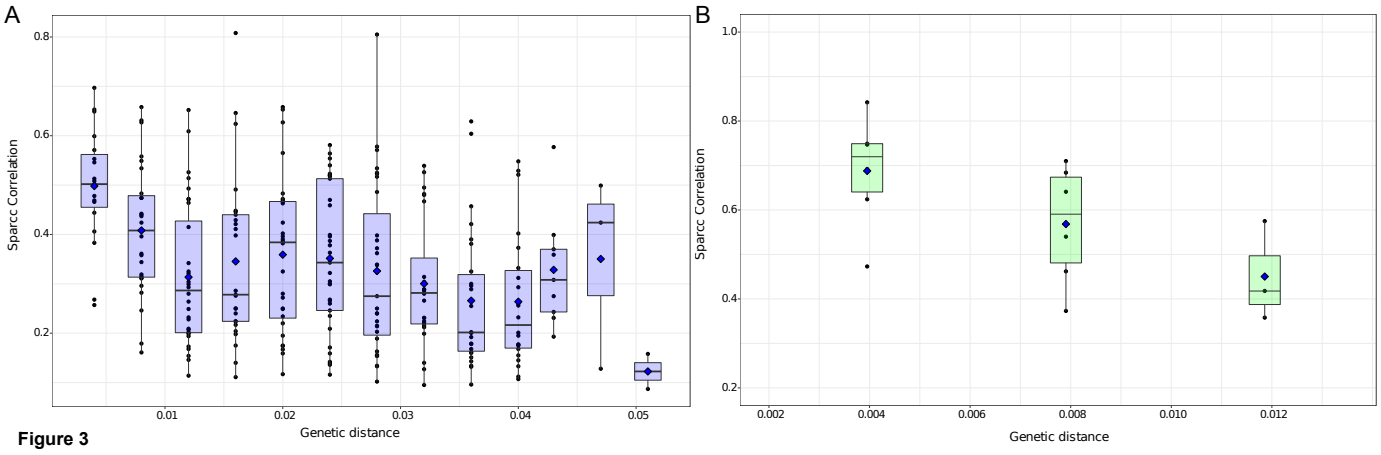
306

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. Relative abundances of each MED node (strain) was computde 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.



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

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



340

341 **Figure 3. Co-occurrence of strains declines with their pairwise genetic distance.** Relationship
342 between co-occurrence (significant SparCC correlation, $P < 0.05$) and genetic distance (p
343 distance) between *Dolichospermum* (A) and *Microcystis* (B) nodes. Blue diamonds represents the
344 mean SparCC correlation for each distance. Boxplots show the median (horizontal line), the 25th
345 and 75th percentile (enclosed in box) and 95% confidence intervals (whiskers).

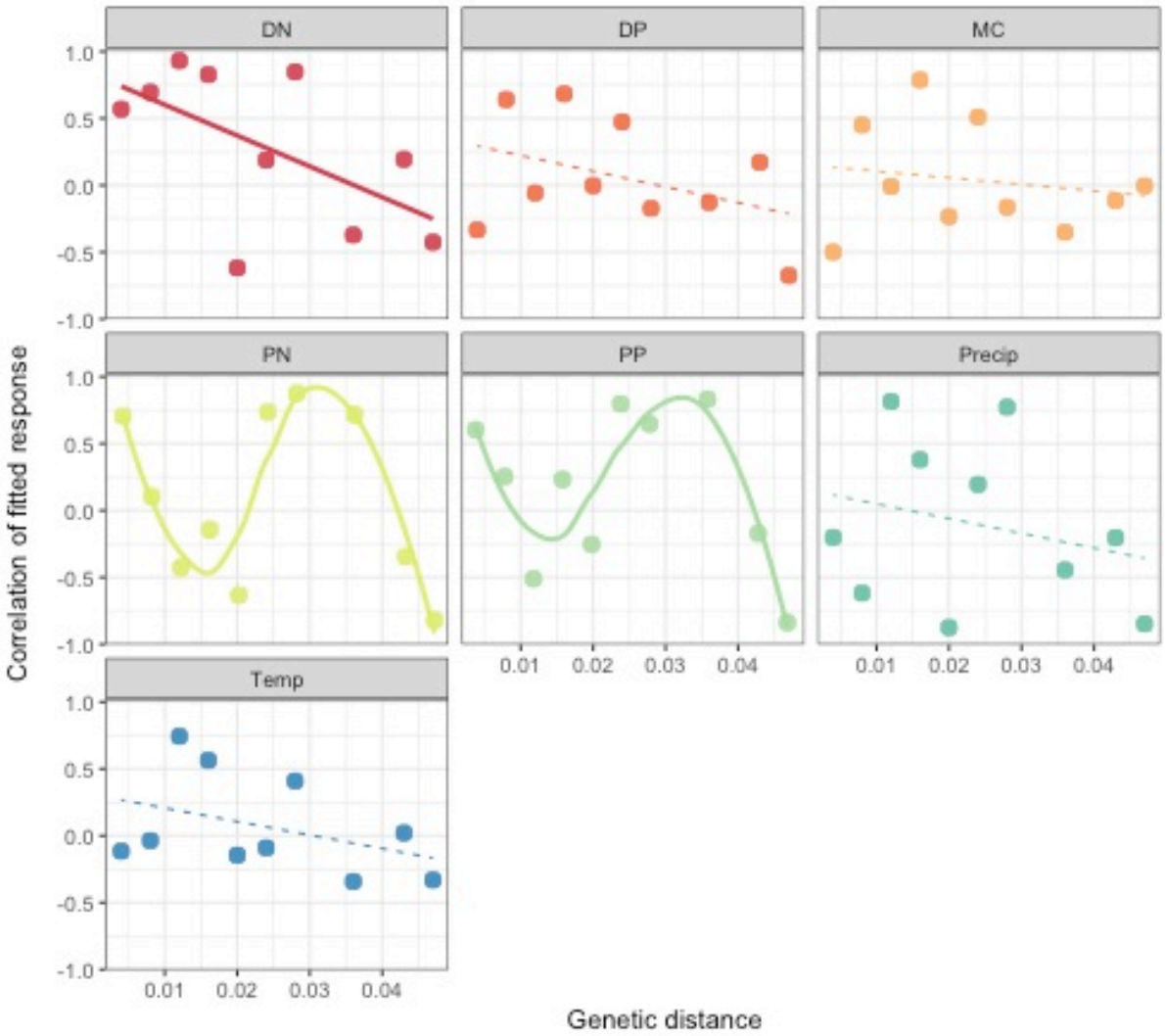
346

347 This pattern was significant within both *D* (linear regression, $F(1,279) = 28.3$, $P < 0.001$, adjusted
348 $R^2 = 8.9\%$) and *M* (linear regression, $F(1,13) = 7.9$, $P < 0.05$, adjusted $R^2 = 33.0\%$). The higher
349 R^2 observed for *M* might be explained by its more limited genetic diversity (maximum pairwise
350 genetic distance ~0.01) compared to *D* (maximum distance ~0.05). There appears to be a rapid
351 decline in niche similarity as genetic distance goes from 0 to ~0.01 (adjusted $R^2 = 23.0\%$ in *D*

352 when considering only this distance range) followed by a flatter relationship for genetic distance
353 >0.01.

354 We also identified non-cyanobacterial MED nodes that co-occurred with each *D* and *M*
355 node, as a proxy for microbial interactions that define their biotic niches. We found that *M* nodes
356 were generally more connected with other members of the bacterial community (Figure S9), and
357 relatively few taxa (5 out of 26; Table S5) are correlated with both *M* and *D*. Overall, this
358 suggests that *M* and *D* strains co-occur with distinct sets of other bacteria, and thus have distinct
359 realized biotic niches.

360 Finally, we asked whether the decline in niche similarity with genetic distance (Figure 3)
361 was a common feature of all niche dimensions, or if different abiotic parameters showed different
362 patterns. We considered all significant models, both linear and non-linear as shown in Figure 4
363 (for *D*) and Figure S8 (for *M*). We found a negative linear relationship between DN niche
364 similarity and genetic distance (adjusted R^2=25%, P = 0.0791). Qualitatively, most other niche
365 dimensions showed a similar pattern for *D* (Figure 4), and among *M* nodes we found a significant
366 negative linear relationship between the correlated fitted response to microcystin concentrations
367 and genetic distance (Figure S8). However, we also identified non-linear relationships between *D*
368 niche similarity and genetic distance for PP (adjusted R^2=69%, P = 0.0184) and PN (adjusted
369 R^2=63%, P= 0.0280). For these niche dimensions, the correlation of fitted responses declines
370 from genetic distances of 0 to ~0.01, then rises to another peak around ~0.03 before declining
371 again (Figure 4).



372

373 **Figure 4. The relationship between genetic distance the co-response (niche similarity) for**
374 ***Dolichospermum*.** LVMs were used to identify correlations between the responses of MED nodes
375 to each measured environmental parameter (separate panels). Positive correlations of fitted
376 responses indicate similar niches; negative correlations indicate different niche preferences.
377 Genetic distances were computed using the p-distance. Separate model fits were tested for the
378 relationship between each niche dimension and genetic distance. See Table S4 for details of
379 model fits.

380

381

382 **Discussion**

383 Overall, our results show how *Microcystis* and *Dolichospermum* can achieve similar levels of
384 dominance during cyanobacterial blooms by partitioning niche space at the genus and sub-genus
385 level. Over eight years of sampling, we observed that *Dolichospermum* and *Microcystis* generally
386 co-occurred during blooms, suggesting a broadly similar physico-temporal niche (Figure 1;
387 Figure S1). However, we confirmed that *D* was associated with lower concentrations of dissolved
388 nitrogen, consistent with its known ability to fix nitrogen (Harke et al., 2016, Andersson et al.,
389 2015) while *M* was associated with higher concentrations of dissolved phosphorus (Figure S5).

390 To dissect niche preferences at finer taxonomic resolution, below the genus level, we used
391 minimum entropy decomposition (MED), allowing single-nucleotide resolution of 16S amplicon
392 sequences. Finer taxonomic resolution has been shown to increase the power to correctly identify
393 phylogenetic overdispersion, a signature of competitive interactions being more important than
394 habitat filtering (Koeppel and Wu, 2014). However, even at high resolution, the 16S marker gene
395 may be too slow-evolving to be a good marker for fast-evolving traits, such as toxin production
396 (Berry et al., 2017) – a trait which likely evolves rapidly by horizontal gene transfer (Moffit and
397 Neilan, 2004).

398 Despite these limitations, MED analysis provided additional insights into the distinct
399 niches of *M* and *D*. *Dolichospermum* diversity was higher and most of the MED nodes (strains)
400 were conditionally rare, some of them being bloom-associated in some years but not in others
401 (e.g D2282, Figure S3). *Microcystis* nodes, on the other hand, were less diverse but more
402 consistently bloom-associated (nodes M5732 and M5733; Figure S4) and more correlated with
403 other taxa (Figure S9), suggesting that the two dominant bloom-forming genera might use
404 different ecological strategies, beyond what is already known about nitrogen and phosphorus
405 utilization.

406 Examining niche partitioning below the genus level also allowed us to detect patterns not
407 evident at the genus level. For example, *Dolichospermum* is associated with higher
408 concentrations of PP and PN at the genus level (Figure S5), but contains a strain (D5630) that
409 shows the opposite pattern (Figure 2). Within *Microcystis*, niche partitioning occurred mainly for
410 temperature and DN, and the two most dominant *M* nodes had a significant niche separation for
411 DN (Figure 2; Figure S7). It is known that *Microcystis* does not fix atmospheric nitrogen, but is
412 able to use refractory N-containing compounds such as urea or amino acids (Moisander et al.,
413 2009; Dai et al., 2009), and DN is likely needed for toxin production (Monchamp et al., 2014).
414 Therefore, different *M* strains could specialize in their preference for different forms of nitrogen.

415 The relative importance of habitat filtering and competition in shaping microbial
416 communities is widely debated, and distinguishing between the two processes can be technically
417 challenging (Koeppel and Wu, 2014; Cadotte and Tucker 2017). Consistent with a general effect
418 of habitat filtering in selecting for genetically similar cyanobacteria under similar conditions, we
419 found a negative relationship between MED node co-occurrence and pairwise genetic distance
420 (Figure 3). This result is in agreement with an early study showing the importance of habitat
421 filtering in microbial communities (Horner-Devine and Bohannan, 2006) and is also consistent
422 with what was observed by Silverman et al (2017) using human microbiome data.

423 Also consistent with the general importance of habitat filtering in shaping cyanobacterial
424 communities, we observed that closely-related strains tended to have similar co-responses to
425 several measured environmental parameters. For example, we found negative linear relationships
426 between niche similarity and *Dolichospermum* genetic distance for most niche dimensions,
427 particularly DN (Figure 4). In *Microcystis,* we found that more genetically similar strains tended
428 to be observed at more similar concentrations of the toxin microcystin (Figure S8). Although 16S
429 is a poor marker for microcystin production (Berry et al., 2017), our results suggest that

430 phylogenetic relatedness is nevertheless somewhat predictive of microcystin concentrations. This
431 means that microcystin production or tolerance is more likely to be shared by close relatives.

432 In contrast to the general pattern of declining niche similarity with genetic distance, PN
433 and PP both had non-linear relationships with genetic distance in *Dolichospermum* (Figure 4).
434 After an initial decline in co-responses to PN and PP to a genetic distance of ~0.01, co-response
435 rose to another peak at ~0.03 before declining again. This response is consistent with near-
436 identical MED nodes sharing the same preferences for PN and PP concentrations, and that
437 competition between close-relatives (up to genetic distance of ~0.01) imposes divergent selection
438 for distinct niche preferences. The similarity in PN and PP niches for more distant relatives
439 (distance ~0.03) can be explained if distant relatives have diverged to reduce competition in other
440 niche dimensions, allowing them to converge in PN and PP preferences. It is unclear why this
441 non-linear relationship is observed for PN and PP, but not for other niche dimensions. However,
442 it highlights the importance of considering each niche dimension separately, because adaptation
443 to different niche dimensions can occur at dramatically different rates (Martiny et al., 2015).

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

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. *Estuarine, Coastal and Shelf Science* 164, 161–171. doi:10.1016/j.ecss.2015.07.013.

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.

Beversdorf, L. J., Miller, T. R., and McMahon, K. D. (2013). The Role of Nitrogen Fixation in Cyanobacterial Bloom Toxicity in a Temperate, Eutrophic Lake. *PLOS ONE* 8, e56103. doi:10.1371/journal.pone.0056103.

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. *Ecology Letters* 12, 693–715. doi:10.1111/j.1461-0248.2009.01314.x.

Crump, B. C., Kling, G. W., Bahr, M., and Hobbie, J. E. (2003). Bacterioplankton Community Shifts in an Arctic Lake Correlate with Seasonal Changes in Organic Matter Source. *Appl. Environ. Microbiol.* 69, 2253–2268. doi:10.1128/AEM.69.4.2253-2268.2003.

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.

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. *The ISME Journal* 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

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.

Höfle, M. G., Haas, H., and Dominik, K. (1999). Seasonal dynamics of bacterioplankton community structure in a eutrophic lake as determined by 5S rRNA analysis. *Appl. Environ. Microbiol.* 65, 3164–3174.

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.

Lindström, (2000). Bacterioplankton Community Composition in Five Lakes Differing in Trophic Status and Humic Content. *Microb. Ecol.* 40, 104–113.

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. *Applied and Environmental Microbiology* 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.

Morrison, J. M., Baker, K. D., Zamor, R. M., Nikolai, S., Elshahed, M. S., and Youssef, N. H. (2017). Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLoS One* 12. doi:10.1371/journal.pone.0177488.

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.

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

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

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

Rohwer, R. R., Hamilton, J. J., Newton, R. J., and McMahon, K. D. (2017). TaxAss: Leveraging Custom Databases Achieves Fine-Scale Taxonomic Resolution. *bioRxiv*, 214288. doi:10.1101/214288.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi:10.1128/AEM.01541-09.

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., et al. (2014). Conditionally Rare Taxa Disproportionately Contribute to Temporal Changes in Microbial Diversity. *mBio* 5, e01371-14. doi:10.1128/mBio.01371-14.

Shade, A., Kent, A. D., Jones, S. E., Newton, R. J., Triplett, E. W., and McMahon, K. D. (2007). Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol. Oceanogr.* 52, 487–494. doi:10.4319/lo.2007.52.2.0487.

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

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

Šimek, K., Kasalický, V., Horňák, K., Hahn, M. W., and Weinbauer, M. G. (2010). Assessing Niche Separation among Coexisting Limnohabitans Strains through Interactions with a Competitor, Viruses, and a Bacterivore. *Appl. Environ. Microbiol.* 76, 3762–3762. 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 Cells1. *Journal of Phycology* 45, 1003–1009. doi:10.1111/j.1529-8817.2009.00723.x.

586 **Data availability**

587 Raw sequence data have been deposited NCBI GenBank under BioProject number

588 PRJNA353865: SRP099259.

589 **Conflict of Interest**

590 The authors declare no conflict of interest.

591 **Acknowledgments**

592 We thank Francis Hui, Andrew Letten and David Warton for advice on the LVM analyses.

593 We also thank everyone who participated in sampling, data collection and analysis.

594 This research was funded by a Natural Sciences and Engineering Research Council (NSERC)

595 Discovery grant and a Fonds de Recherche du Québec Nature et Technologies (FRQNT) New

596 Researcher grant to BJS, and the federal government interdepartmental Genomics Research and

597 Development Initiative (GRDI). NT is funded by a project from the European Union's Horizon

598 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No

599 656647.