The evolution of realized niches within freshwater *Synechococcus*

Authors: Nicolas Tromas[1#]*, Zofia E. Taranu[2#,] Mathieu Castelli[3#], Juliana S. M. Pimentel[4], Daniel A. Pereira[4], Romane Marcoz[1], B. Jesse Shapiro[1*] and Alessandra Giani[4]

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

2- Environnement et Changement Climatique Canada,105 Rue McGill, Montréal, QC H2Y 2E7

3- mathieu.castelli@gmail.com

4- Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

*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

# These authors contributed equally to this work

Originality-Significance Statement

We address a fundamental question in ecology and evolution: how do niche preferences change over evolutionary time? Using time-series analysis of 16S rRNA gene amplicon sequencing data, we develop an approach to highlight the importance of biotic factors in defining realized niches, and show how niche preferences change proportionally with the 16S gene molecular clock within the genus *Synechococcus.* Ours is also one of few studies on the ecology of freshwater *Synechococcus,* adding significantly to our knowledge about this abundant and widespread lineage of Cyanobacteria.

Summary

Understanding how ecological traits have changed over evolutionary time is a fundamental question in biology. Specifically, the extent to which more closely-related organisms share similar ecological preferences due to phylogenetic conservation – or if they are forced apart by competition – is still debated. Here we explored the co-occurrence patterns of freshwater cyanobacteria at the sub-genus level to investigate whether more closely-related taxa share more similar niches, and to what extent these niches were defined by abiotic or biotic variables. We used deep 16S rRNA gene amplicon sequencing and measured several abiotic environmental parameters (nutrients, temperature, etc.) in water samples collected over time and space in Furnas Reservoir, Brazil. We found that relatively more closely-related *Synechococcus* (in the continuous range of 93-100% nucleotide identity in 16S) had an increased tendency to co-occur with one another (*i.e.* had similar realized niches). This tendency could not be easily explained by shared preferences for measured abiotic niche dimensions. Thus, commonly measured abiotic parameters might not be sufficient to characterize, nor to predict community assembly or dynamics. Rather, co-occurrence between *Synechococcus* and the surrounding community (whether or not they represent true biological interactions) may be a more sensitive measure of realized niches. Overall, our results suggest that realized niches are phylogenetically conserved, at least at the sub-genus level and at the resolution of the 16S marker. Determining how these results generalize to other genera and at finer genetic resolution merits further investigation.

Introduction

A bacterial community is a group of potentially interacting organisms that coexist at a particular place and time (Magurran, 2003). Environmental selective pressures are a strong force shaping microbial community assembly (Martiny et al, 2015). We know, for example, that certain abiotic factors explain a large portion of the variation in microbial community composition (e.g. the effect of pH on soil bacterial communities; Fierer and Jackson, 2006). Therefore, associations between microbial traits - generally defined as a phenotypic response to a specific environmental condition - and abiotic niches could help explain community assembly rules (Green et al., 2008; Burke et al., 2011). However, using trait-based approaches to understand communities may be challenging, as important abiotic variables may go unmeasured. Even less is known about biotic interactions, despite their importance in determining community composition, diversity, and dynamics (Needham et al., 2016).

Abiotic factors are generally thought to determine an organism's fundamental niche (where it is theoretically capable of living), whereas biotic factors determine its realized niche (where it actually lives in nature; Hutchinson, 1957). Species (or taxonomic) co-occurrence networks are often used to infer niche similarity among organisms that tend to co-occur in nature over space and time, and microbial co-occurrence networks can easily be constructed from 16S rRNA gene amplicon sequencing surveys (Friedman and Alm, 2012; Röttjers and Faust, 2018; Tromas et al., 2018). However, co-occurrence can be driven by both abiotic and biotic factors, which are hard to disentangle in practice (Kraft et al., 2015). Regardless, organisms sharing a similar niche are expected to be associated with similar surrounding communities (Cohan and Koeppel, 2008; Faust et al.2012; Pascual-García et al., 2014).

80

81 A fundamental question spanning ecology and evolution is how ecological traits change over
82 evolutionary time. For example, some traits (such as bacteriophage host range or phosphorus
83 utilization) evolve rapidly at the tips of a phylogenetic tree, whereas other traits (such as salinity
84 preference) are deeply conserved (Martiny *et al.* 2015). When more closely related organisms
85 share similar ecological preferences, so-called "habitat filtering" or "environmental" filtering is
86 expected to result in phylogenetic clustering, meaning that a community tends to contain more
87 closely related organisms than expected by a random draw from the phylogeny (Webb et al.,
88 2002; Horner-Devine & Bohannan, 2006; Martiny et al., 2015). In contrast, if close relatives
89 evolve different traits to avoid competitive exclusion, this will result in phylogenetic
90 overdispersion (*i.e.* a community composed of more distant relatives than expected by chance).
91 However, the relative importance of these two processes in shaping microbial communities is still
92 widely debated, and difficult to distinguish (Koeppel and Wu, 2014; Cadotte and Tucker, 2017).
93 We have previously shown that within the cyanobacterial genus *Dolichospermum* (Tromas et al.,
94 2018), the relationships between phylogenetic distance and ecological similarity varies by trait,
95 suggesting that it might be necessary to analyze each niche dimension or trait separately (Martiny
96 *et al*., 2015).

97

98 In this study, we explored the co-occurrence patterns of freshwater cyanobacteria at the sub-
99 genus level, to investigate if more closely related taxa have more similar niches, and to what
100 extent these niches can be quantified by abiotic or biotic variables. We use the term "closely
101 related" as a relative term, describing a continuous range of genetic similarity (93-100%
102 nucleotide identity) in 16S rRNA gene sequences. We focused on *Synechococcus*, the most
103 abundant cyanobacterial genus in Furnas Reservoir (Brazil) at the time of sampling (2006-2008).

104 *Synechococcus* is among the most abundant organisms living in oceans and lakes (Stockner et al.
105 2000; Scanlan, 2003). The phylogenetic coherence of the genus has been questioned by a recent
106 study showing it to be polyphyletic (Coutinho et al 2016). *Synechococcus* is physiologically
107 highly plastic, ubiquitous, and able to acclimate to different environmental conditions (Callieri
108 1996; Vörös et al. 1998; Callieri et al., 2011). Previous studies have shown that different
109 *Synechococcus* strains could co-exist in the same site but respond differently to environmental
110 changes, suggesting niche partitioning (Ferris et al., 2003; Allewalt et al., 2006; Becker et al.,
111 2007; Becraft et al., 2011; Callieri et al. 2012). Recently, Zheng et al. (2018), observed a
112 geographical pattern of heterotrophic bacteria associated with different marine *Synechococcus*
113 strains, indicating that strains living in the same area tend to be associated with similar
114 communities. It remains, however, unclear whether the *Synechococcus* genotype or the
115 environment are the main drivers of *Synechococcus* interactions with the surrounding microbial
116 community.
117
118 Using deep 16S rRNA gene amplicon sequencing of 86 water samples collected in time series
119 across nine locations in the Furnas Reservoir, we tracked genetic diversity within the
120 *Synechococcus* genus, along with the surrounding microbial community, and measured several
121 abiotic variables. We found that that more closely-related *Synechococcus* tended to co-occur with
122 one another and also with similar surrounding microbial communities. Such phylogenetic
123 clustering indicates that realized niche similarity tends to evolve proportionally with substitutions
124 in the 16S gene in the *Synechococcus* lineage (*i.e.* niches evolve according to the 16S gene
125 molecular clock). However, more closely-related *Synechococcus* did not have similar abiotic
126 niche preferences (with the exception of total phosphorus). These results suggest that biotic
127 factors may be strong niche determinants, which could be particularly useful when relevant

128 abiotic factors are unknown or cannot be readily measured. Alternatively, cryptic abiotic drivers
129 may determine niche and community structure, but biotic factors provide the most informative
130 measure of niche similarity.
131
132 **Materials and methods**
133
134 ***Sampling, environmental data measurements***
135 A total of 90 water samples were collected from September 2006 to April 2008 at nine stations in
136 Furnas Reservoir (Minas Gerais, Brazil). Furnas is a large reservoir (1440 km^2), located in
137 southeastern Brazil (20º40'S; 46º19'W) and formed by the damming of two main rivers (Rio
138 Grande and Rio Sapucaí), which divide the reservoir in two separated branches (Figure S1).
139 Sampling stations F1, F4, F6, F9 are from a relatively pristine branch of the reservoir, whereas
140 F12, F14, F18 and F20 are impacted by human activities. Temperature profiles were measured in
141 the water column by aid of a multi-parameter probe (556 YSI, USA). Water samples were
142 collected from the euphotic zone (determined by Secchi disc depth) by a Van Dorn vertical water
143 sampler. Samples were stored in bottles that had been acid-washed and rinsed with deionized
144 water. A portion of each water sample was immediately filtered through glass fiber filters
145 (Whatmann GF/F, 0.7 μm pore size). The exact filtered volumes were recorded and filters were
146 kept frozen until further analyses (chlorophyll a, DNA and microcystin). For dissolved nutrient
147 analyses, 200 mL of filtered water samples were stored at −20 °C. For total phosphorus, samples
148 were frozen with no previous filtration. Nutrient analyses were performed using
149 spectrophotometric methods according to APHA (2005). All nutrient analyses (Nitrate, Nitrite
150 and total phosphorus (TP) were performed on three replicates.
151
152 ***DNA extraction, purification and sequencing***

153 DNA was extracted from frozen filters according to Kurmayer et al. (2003), with few
154 modifications. Briefly, filters were treated with a sucrose buffer (25% w/v sucrose, 50 mM Tris-
155 HCl, 100 Mm EDTA, pH 8) on ice for 2 h and with addition of lysozyme (5mg/mL, 1h, 37°C).
156 Proteinase K (100 µg/mL) in sodium dodecyl sulfate (2% v/v) was added and filters were
157 incubated overnight at 55°C. A phenol:chloroform:isoamyl alcohol solution (25:24:1, v/v) was
158 used for protein precipitation and DNA isolation. The DNA was cleaned in 100% ethanol and
159 pellets rinsed with 70% ethanol. The DNA was resuspended in TE (10 mM Tris- HCl, pH 8, and
160 1 mM EDTA). The DNA extract was quantified by a spectrophotometer (Perkin Elmer, Lambda
161 25), at 260 nm and 280 nm, and its quality checked in 1% (w/v) agarose gel, stained with
162 ethidium bromide. DNA samples were stored at -20ºC.
163
164 ***Sequence analysis***
165 We followed the same protocol described in Tromas et al., (2017) to generate a library of
166 V4 region amplicons. Libraries for 2x250bp paired-end Illumina sequencing were prepared using
167 a two-step 16S rRNA gene amplicon PCR as described in Preheim et al. (2013). We performed
168 one sequencing run using MiSeq reagent Kit V2 (Illumina, San Diego, CA, USA) on a MiSeq
169 instrument (Illumina). A total of 4,476,747 sequences of the 16S rRNA gene V4 region were
170 obtained from 90 lake samples, two negative controls, and two mock community samples. We
171 obtained a median of 37,682 sequences per sample. Using a similar approach as described in
172 Tromas *et al*., (2018), we processed the sequences with SmileTrain
173 (https://github.com/almlab/SmileTrain/wiki; Preheim et al., 2013) for read quality filtering,
174 primer removal, chimera filtering, and merging using USEARCH (version 7.0.1090,
175 http://www.drive5.com/usearch/, default parameter) (Edgar, 2010), Mothur (version 1.33.3)
176 (Schloss *et al*., 2009), and Biopython (version 2.7). Minimum Entropy Decomposition (MED)

177 was then applied to the filtered and merged reads to partition sequence reads into MED nodes
178 (Eren *et al*., 2015). MED was performed using the following parameters: –M noise filter set to
179 1000, resulting in ~17.5% of reads filtered and 466 MED nodes representing the whole bacterial
180 community. Samples with less than 1000 reads were removed, yielding a final dataset of 86
181 reservoir samples. Finally, we assigned taxonomy to MED nodes using the assign_taxonomy.py
182 QIIME script (default parameters), and a combination of GreenGenes and a freshwater-specific
183 database (Freshwater database 2016 August 18 release; Newton *et al*., 2011), using TaxAss
184 (https://github.com/McMahonLab/TaxAss, installation date: September 13th 2016; Rohwer *et al*.,
185 2017). We found a total of 40 *Synechococcus* nodes, which we refer to as strains below.
186
187 ***Spatio-temporal analysis***
188 We used multivariate regression tree analyses (Breiman *et al*. 1984; De'ath 2002) to
189 investigate if spatio-temporal variables could explain genetic variation within *Synechococcus*. We
190 used two different temporal predictors: year and month and one spatial predictor: station. The
191 analysis was performed using the function *mvpart*() and *rpart.pca*() from the R mvpart package
192 (Therneau and Atkinson, 1997; De'ath, 2007). Prior to analysis, the *Synechococcus* MED node
193 table was Hellinger transformed to downweight the effect of double-zeros (Rao, 1995).
194
195 ***Genetic distance between Synechococcus nodes***
196 We measured the genetic distance between *Synechococcus* MED nodes using the software
197 MEGA (Kumar et al., 2016; version 7.0.18) with the p-distance (the proportion of nucleotide
198 sites at which two sequences differ), calculated by dividing the number of sites with nucleotide
199 differences by the total number of sites compared (excluding sites with gaps).
200

201 ***Co-response to abiotic factors***

202 As described in Tromas *et al.* (2018), we used a Latent Variable Model (LVM)
203 framework (boral package in R; Hui 2015, Warton *et al.* 2015) to explore how *Synechococcus*
204 nodes co-responded to abiotic gradients and used these co-responses as indicators of niche
205 similarity. That is, for each abiotic factor, we ran separate LVMs, regressing the bacterial
206 community as a function (both linear and non-linear) of the given factor. This component of the
207 LVM thus defined the taxon-specific environmental responses. To then identify remaining
208 patterns of co-occurrence after accounting for all measured environmental variables, we fit a
209 global LVM which included all abiotic factors and two latent variables (*sensu* Letten *et al.* 2015
210 and Warton *et al.* 2015). To visualize patterns of co-occurrence arising from the different
211 environmental factors, we calculated two types of correlation matrices. The first, a co-response
212 correlation matrix, was constructed by calculating, for any two nodes, the correlation between
213 their fitted values. This correlation matrix thus represented the correlation between nodes that can
214 be attributed to a shared or diverging environmental responses. A significant positive correlation
215 between the fitted response of any two taxa to an environmental variable represented a co-
216 response (i.e., as one taxa increases in response to an environmental variable, the other likewise
217 increases), whereas a significant negative correlations between the fitted response of two taxa
218 represented some degree of niche separation (i.e., as one taxa increases in response to an
219 environmental variable, the other decreases).

220

221 To account for the correlation between nodes that may be attributable to biotic processes or
222 missing environmental covariates, a second type of correlation matrix, a residual correlation
223 matrix, was calculated using the latent variable coefficients of the global LVM. Since the latent
224 variables are the output of an ordination of the residuals, we expected weak residual correlations

225 induced by the latent variables if the environmental variables were the dominant force structuring
226 patterns of species co-occurrence (*i.e.* the model explained most of the variance, with little left
227 unexplained in the residuals). Conversely, if any unmeasured environmental factors and/or biotic
228 processes are equally, or more, important than measured environmental factors, we expected
229 strong correlations based on the latent variables (*i.e.* the model was poor and much of the
230 observed variability remained in the residuals).

231 To test how *Synechococcus* niche similarity varied with genetic distance, we examined the
232 relationship between the co-response of *Synechococcus* taxa to environmental parameters and
233 their genetic distance (i.e., plotting the correlation coefficient of the LVM co-responses vs.
234 genetic distances) (R_script1). However, given the large number of environmental factors driving
235 phytoplankton community dynamics (Hutchinson, 1961), we expected some degree of co-
236 limitation whereby the response of a taxon to an environmental variable (and consequently the
237 degree of co-response among any two taxa) would be limited by other environmental variables.
238 In such cases, the taxon's response and rate of change would have an upper limit (set by all
239 measured environmental factors) but may not reach this limit if other, unmeasured factors are co-
240 limiting (Cade and Noon 2003). As an increasing number of unmeasured factors become limiting
241 at some sample location, or time point, the relationship between the response and the measured
242 factor becomes increasingly heterogeneous or wedge shaped. When such heterogeneous
243 variances are observed (wedge shape biplot), it suggests that there is not a single slope coefficient
244 that characterizes the relationship, and that focusing solely the coefficient fit by ordinary least
245 squares regression (mean response) may underestimate the true rate of change. Thus, to
246 accurately examine co-limitations among measured and unmeasured factors, we applied quantile
247 regressions using 'quantreg' R package (Koenker, 2015) (R_script1), which, instead of fitting a
248 regression to the mean response (ordinary least squares regression), fits regression curves to other

249 quantiles of the response variable (Cade and Noon 2003). Finally, to provide an effect size for the
250 linear quantile relationships of abiotic co-responses vs. genetic distance, we calculated the
251 goodness of fit measure for quantile regression for a suite of quantiles (where the goodness of fit
252 is estimated as 1 minus the ratio between the sum of absolute deviations of the parameterized
253 model and the sum of absolute deviations of the unparameterized, null model). Note that this
254 goodness of fit measure (referred hereafter as R1) is not comparable to the OLS coefficient of
255 determination (R2); it is based on the variance of absolute deviations as opposed to the variance
256 of squared deviations, and consequently, R1 is always smaller than R2 (Koenker and Machado,
257 1999).
258
259 ***Co-occurrence with biotic factors***

260 A caveat of the above LVM-quantile regressions is that although it identifies whether
261 some relationships are limited by unmeasured abiotic and/or biotic factors, it could not tease apart
262 the relative importance of each type of factor (everything is instead lumped as an unknown). In
263 order to determine the relative contribution of biotic processes, we thus quantified node co-
264 occurrence patterns among *Synechococcus* and the remaining community. In particular, we
265 explored the relationship between *Synechococcus* nodes and the surrounding community by
266 measuring: (1) co-occurrences between *Synechococcus* and other taxa, and (2) paired differences
267 in co-occurrence among *Synechococcus* nodes and a specific taxon. For the former, we calculated
268 co-occurrences among taxa using SparCC (Friedman and Alm, 2012), including 20 iterations to
269 estimate the median correlation of each pair of MED nodes, 500 bootstraps to assess the
270 statistical significance and centered log ratio (CLR) transform to correct for compositionality.
271 Correlations were then filtered using a false discovery rate threshold ($Q < 0.05$). For the latter, we
272 further calculated the absolute difference of SparCC correlation (r) between a *Synechococcus*

node (X_i) and a specific taxon T, such that $|\Delta r| = |\ Corr(X_1, T) - Corr(X_2, T)\ |$ where Corr is defined here as the SparCC correlation between X_i and T.

For each non-*Synechococcus* taxon, we then estimated the relationship between $|\Delta r|$ and the distance between the given *Synechococcus* nodes (X_1 and X_2). A positive correlation would indicate that more closely related *Synechococcus* nodes have a lower $|\Delta r|$ than more distant nodes, *i.e.* more closely related nodes would have more similar correlations with potentially interacting community members. To reduce potential bias or noise due to a small sample size of *Synechococcus* pairs, we selected all non-*Synechococcus* nodes that were significantly co-occurrent with at least seven different *Synechococcus*. This cutoff was chosen to ensure at least 20 unique pairs of *Synechococcus* for each non-*Synechococcus* taxon. A total of 373 non-*Synechococcus* taxa were selected using this cutoff, and the correlation between $|\Delta r|$ and genetic distance was non-significant (Spearman correlation, $P<0.05$) for 120 of them. Finally, we performed a permutation test to quantify any bias in the method due to data structure. For each of the significant non-*Synechococcus* taxa, we estimated the rate of false positive correlations by sampling values of *Synechococcus* genetic distance with replacement, while randomizing the association between genetic distance and $|\Delta r|$. We performed 1000 permutations for each significant non-*Synechococcus* taxon using the script R_script2. We then recorded the proportion (p) of permutations yielding a larger correlation than observed, after adding a pseudocount of 1 to both the numerator (the number of permutations yielding a larger correlation than observed) and the denominator (the total number of permutations).

Phylogenetic analysis

We first verified that our *Synechococcus* sequences were monophyletic by building a phylogenetic tree with FastTree (version 2.1.8, Price et al., 2009), using all MED node sequences

298 aligned with MAFFT (v7.154b; Katoh and Standley, 2013). We then specifically aligned
299 *Synechococcus* MED node sequences using muscle (in Mega, version 7.0.26; Kumar et al., 2016)
300 and built phylogenetic trees using PhyML with the general time-reversible (GTR) model of
301 nucleotide substitution (version 3.0; Guindon et al., 2010). We used the ALDEx2 R package
302 (version: 1.5.0; Fernandes *et al*., 2014) and the aldex() function to identify MED nodes associated
303 with the pristine or human-impacted branches of Furnas Reservoir. To do so, we used Welsh's t-
304 test and 128 Monte Carlo samples. ALDEx2 uses the CLR transformation to avoid compositional
305 effects. A Q-value below 0.05 after Benjamini-Hochberg correction was taken as evidence for
306 association. We also used a second association method, DESeq2 implemented in the
307 "MicrobiomeAnalyst" web-based tool (Dhariwal et al., 2017).
308
309 ***Statistical analysis***
310
311 All statistical tests were performed in R.
312
313

314 **Results**

315

316 *Abiotic factors shape Synechococcus community structure*

317 Furnas Reservoir is divided into two branches: one, which is heavily human-impacted (Figure S1,
318 sampling stations 12-20) and the other, less impacted (stations 1-9). *Synechococcus* dominated
319 the bacterial community in both branches during the sampling period (2006-2008) (Figure S2)
320 and formed a monophyletic group (Figure S3, S4). Furnas Reservoir is meso-eutrophic based on
321 phosphorus concentrations. Total phosphorus (TP) did not vary significantly across stations but
322 did vary significantly within each station over time (Figure S5). Temperature did not vary widely
323 across sampling stations or time. We first asked whether particular *Synechococcus* strains (which
324 we used interchangeably with MED nodes) were associated with different branches of the
325 reservoir, different time periods, or different abiotic factors. We applied a multivariate regression
326 tree analysis (Breiman et al. 1984; De'ath 2002) to test whether the composition of
327 *Synechococcus* strains changed over time and space. We found that the year and month of
328 sampling did not explain much variance (R^2=0.02 and R^2=0.03, respectively), nor did the station
329 (R^2=0.06) or branch of the reservoir (R^2=0.02). In contrast, Nitrite, TP, depth, and water
330 temperature all explained a significant proportion of the variance (Figure S6; 22.5% of variance
331 explained). Nitrite is the single most explanatory variable (13% of variance, but also colinear
332 with temperature and depth; Figure S6), followed by TP (5.5% of variance; Figure S7).

333

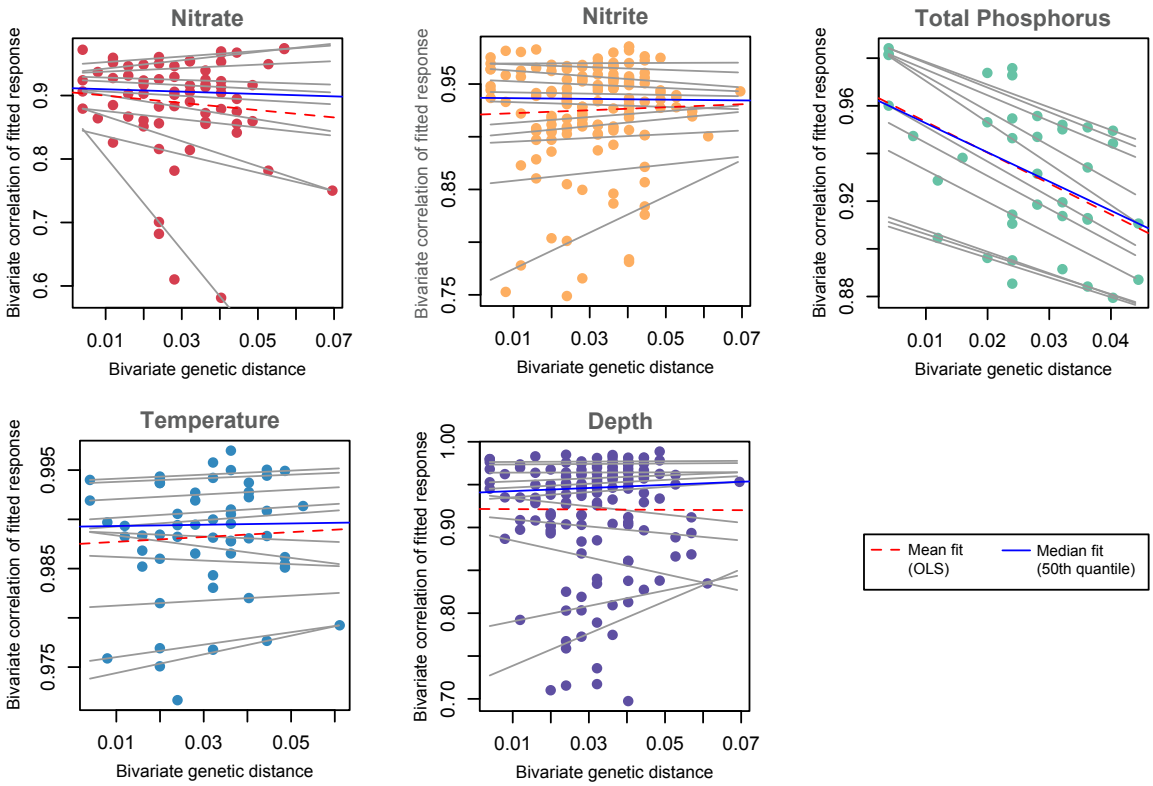
334 *More closely related Synechoccus do not share similar preferences for most measured abiotic*
335 *niches*

336 Having established that abiotic factors, but not temporal or spatial variation, explain a significant
337 amount of variation in the relative abundances of *Synechococcus* strains, we asked whether more

338 closely related strains tended to share similar abiotic niches. We examined *Synechococcus* 16S
339 sequences across a range of ~93-100% nucleotide identity, which is a rather large range
340 considering that niche partitioning can occur at relatively high identity, >97% (Martiny *et al.*
341 2015 ; Larkin and Martiny, 2017). Here we defined abiotic niche similarity by estimating the co-
342 responses of pairs of *Synechococcus* strains to each abiotic factor using LVM-quantile regression
343 (Methods). For most abiotic factors, we did not observe any significant relationship between co-
344 responses and genetic distance, with the exception of total phosphorus for which the relationship
345 was negative (Median p-value across quantiles = 0.021; median R1 across quantiles = 0.12)
346 (Figure 1 & S8, Table S2). Similarly, the degree of abiotic niche separation (*i.e.* significant
347 negative correlations between the fitted response of any two *Synechococcus* strains to an abiotic
348 factor) was generally independent of genetic distance (Figure S9, S10), except for depth and
349 Nitrite which were significant at some quantiles but with quite a poor fit (median R1 across
350 quantiles ≤ 0.02)(Table S2C).

351 We also evaluated non-linear quantile relationships among co-responses and genetic
352 distances, but we never found a significant non-linear relationship over the total range of genetic
353 distances (Table S2A-C). We repeated the quantile regression analysis in a restricted range of
354 100-97% identity, to see if any linear trend is discernible in a range roughly corresponding with
355 species or genera (Figure S11). The negative linear trend observed between co-responses and
356 genetic distance for TP in the full range of genetic identify (Figure 1) was still observed in the
357 restricted range (Figure S11) although not significant at all quantiles (Table S2D). There was also
358 some evidence for a non-linear relationship between genetic distance and both positive and
359 negative co-responses to TP (Table S2 D, E), suggesting possible niche separation among closely
360 related (~99% identical) strains. However, this trend is driven by just three pairs of very closely
361 related strains with quite different TP preferences (Figure S12). Overall, these results suggest that

362 TP-related niches show some evidence for phylogenetic conservation within freshwater
363 *Synechoccus*, while the evidence for other niches is weak or inconclusive.



364

365 **Figure 1. Relationship between *Synechococcus* niche similarity and genetic distance.** Niche
366 similarity is defined as a positive LVM co-response (y-axis). Shown are the scatterplots of LVM
367 co-responses versus each variables superimposed with the lines for 5th, 10th, 15th, 25th, 35th, 45th,
368 55th, 65th, 75th, 85th and 90th quantile regression fits (grey lines, for each co-response quantile),
369 the median fit (50th quantile; blue line), and the least squares estimate or mean fit (dashed red
370 line).
371

372 Biotic processes or unmeasured environmental variables could define *Synechococcus* realized
373 niches, and are captured in the residual variation of the global LVM. We next asked whether
374 more closely-related *Synechococcus* strains had more similar co-responses to this residual
375 variation. For the positive co-responses (Figure 2A, B), we observed a decrease with genetic

376 distance. This result suggests a negative relationship between the residuals (*i.e.* unmeasured
377 abiotic variables and/or biotic niche preferences) and phylogenetic distance. For the negative
378 residual correlations (Figure 2C, D), there was no apparent relationship with genetic distance.
379 Therefore, there remain significant co-responses that were not explained by the abiotic factors
380 measured in this study. Overall, most of the measured abiotic environmental variables were not
381 related to *Synechococcus* genetic distance. However, preferences for unmeasured biotic or abiotic
382 factors do tend to change with genetic distance, such that more closely-related *Synechococcus*
383 have similar realized niche preferences.
384

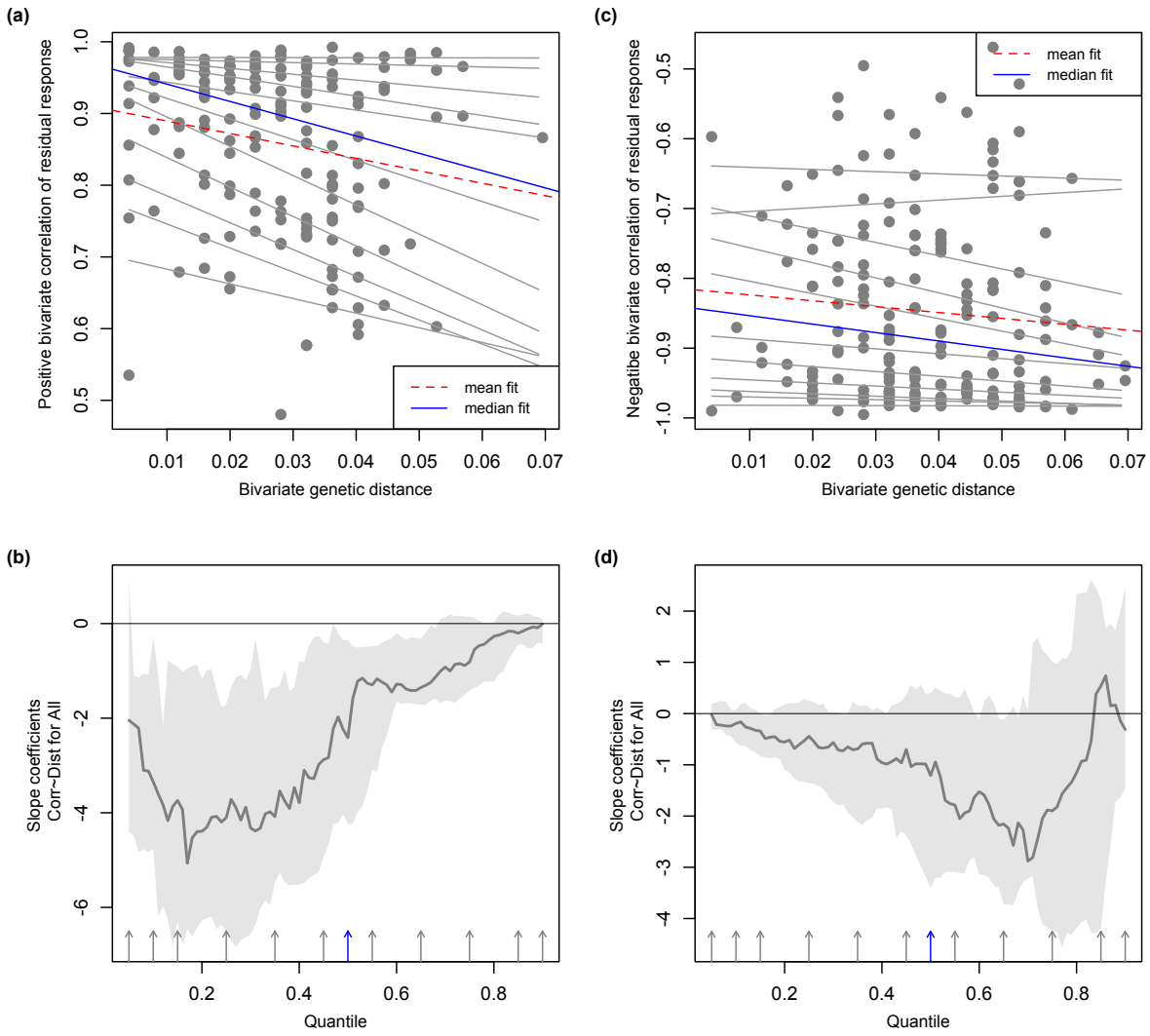
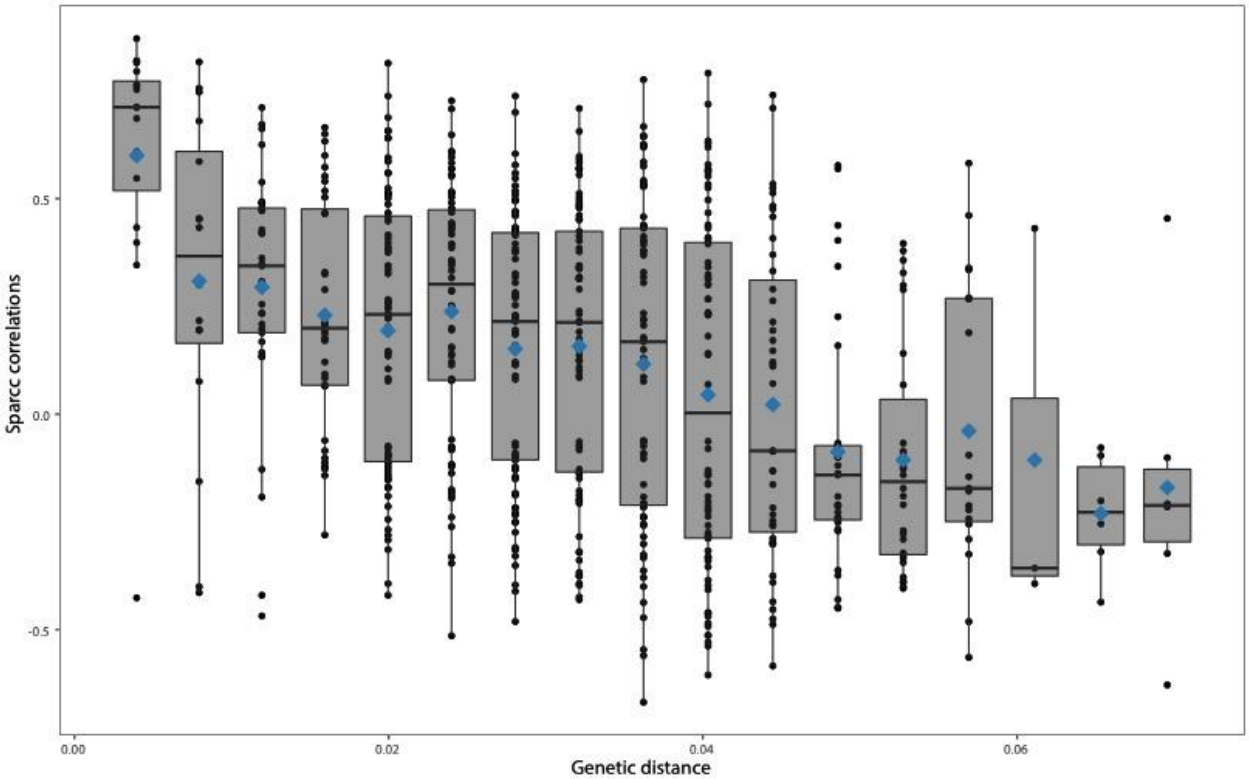


Figure 2. Relationship between *Synechococcus* co-responses (LVM residuals) and genetic distance. Positive co-responses are shown in panels **A** and **B**; negative co-responses in panels **C** and **D.** *(*A, C*)* Scatterplots of LVM co-responses versus LVM residuals superimposed with the lines for 5th, 10th, 15th, 25th, 35th, 45th, 55th, 65th, 75th, 85th and 90th quantile regression fits (grey lines), the median fit (50th quantile; blue line), and the least squares estimate or mean fit (dashed red line). (B, D) Panels show the quantile slope estimates (dark grey line) and corresponding confidence intervals (light grey bands) of the response models across all quantiles (arrows, with the median shown in blue). Significant quantile slopes occur when confidence intervals do not overlap zero.

396 *More closely related Synechococcus share similar biotic interactions*

397 The observed tendency for more closely-related *Synechococcus* to share similar realized niches
398 (Figure 2A) could be explained by shared preferences for either unmeasured abiotic factors or
399 biotic interactions. We used co-occurrence network analysis to explore the role of potential biotic
400 interactions. First, we confirmed that more closely related *Synechococcus* tend to co-occur across
401 samples. We observed a negative relationship between node co-occurrence and pairwise genetic
402 distance (Figure 3), indicating that genetically similar *Synechococcus* nodes are indeed more
403 likely to co-occur, and thus to have similar realized niches. This pattern was significant [linear
404 regression, $F(1,488) = 108.6$, $P < 0.001$, adjusted $R^2 = 0.18$].

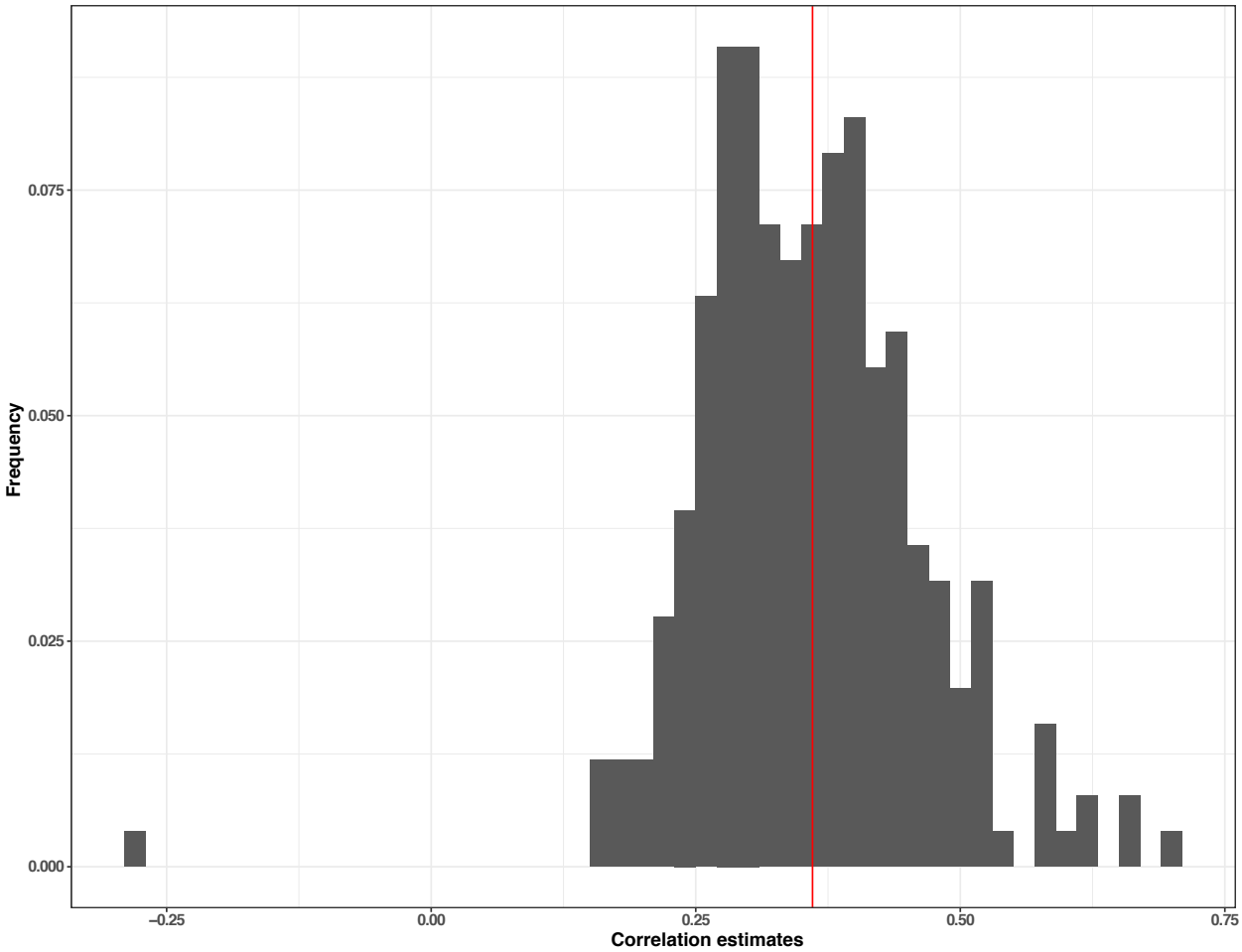
405



406

407 **Figure 3**. **Negative relationship between pairwise SparCC correlation coefficients and**
408 **pairwise genetic distance between *Synechococcus* nodes.** Only significant SparCC correlations
409 ($Q < 0.05$) were included. Pairwise genetic distance was computed as percent nucleotide identity

410 (p-distance). Blue diamonds represent the mean SparCC correlation for each distance. Boxplots
411 show the median (horizontal line), the 25th and 75th percentile (enclosed in box) and 95%
412 confidence intervals (whiskers).
413
414 If more closely-related *Synechococcus* tend to co-occur, it follows that they should also co-occur
415 with more similar surrounding microbial communities. The composition of the surrounding
416 microbial community could provide information about the nature of the realized niche, while also
417 suggesting possible biotic interactions. To confirm the expectation that more closely-related
418 *Synechococcus* should also share similar surrounding communities, we performed a pairwise
419 analysis to examine the relationship between *Synechococcus* genetic distance and $|\Delta r|$, a measure
420 of the similarity of *Synechococcus* co-occurrence with non-*Synechococcus* taxa (Methods). A
421 total of 373 non-*Synechococcus* taxa were analysed and the correlation between $|\Delta r|$ and genetic
422 distance was non-significant for 120 of them. The remaining 253 significant non-*Synechococcus*
423 have consistently positive relationships between co-occurrence and genetic distance, *i.e.* more
424 closely related *Synechococcus* nodes have a higher chance of co-occurring with similar non-
425 *Synechococcus* taxa (Figure 4). This result is robust to data structure, as determined by permuting
426 the associations between genetic distance and $|\Delta r|$ (Figure S13; Methods).
427
428



429

430 **Figure 4. More closely related *Synechococcus* nodes co-occur with similar surrounding**
431 **community**. The histogram shows the distribution of Spearman correlations between the genetic
432 distance of *Synechococcus* nodes and their association with similar surrounding communities.
433 Spearman correlations were calculated between the genetic distance of *Synechococcus* pairs and
434 the absolute difference of SparCC correlation (r) of a *Synechococcus* node with a specific taxon
435 T, such that $|\Delta r| = |\ Corr(X_1,\ T) - Corr(X_2,\ T)\ |$ where Corr is defined here as the SparCC
436 correlation score of X_i (*Synechococcus* node) and T (non-*Synechococcus* node). We then
437 estimated the correlation of $|\Delta r|$ and the genetic distance between nodes X_1 and X_2. The mean
438 Spearman correlation estimate was 0.373 (red line) and 99.6% of correlations are positive.
439

440 Finally, we selected the non-*Synechococcus* taxa with the highest correlation scores between $|\Delta r|$
441 and *Synechococcus* distance genetic. Five taxa had correlation estimates higher than 0.6,

442 including three Proteobacteria, one Verrucomicrobia, and one Planctomycetes (Table S1).
443 Previous studies have shown that members of the *Comamonadaceae* family and *Spirobacillales*
444 order (members of Proteobacteria; Table S1) were associated with *Synechococcus* (Guedes et al.,
445 2018) or more broadly with algae (Fullbright et al., 2018). Moreover, the *Gallionella* genus was
446 previously observed colonizing filamentous algae (Mori et al., 2015).
447
448 Overall, these results show that the co-occurring surrounding community is, on average, more
449 similar for more closely related *Synechococcus*. Members of this surrounding community could
450 interact directly with *Synechococcus,* or share similarity in their abiotic niches.
451
452 **Discussion**
453 In this study, we investigated *Synechococcus* niche preferences at relatively fine (sub-
454 genus) taxonomic resolution using a 16S sequencing approach previously applied to other
455 bacterial lineages (Koeppel and Wu, 2014; Tromas et al., 2018). However, we acknowledge that
456 even at the finest possible taxonomic resolution, the 16S gene may be too slow-evolving to be a
457 good marker for fast-evolving traits, such as toxin production (Berry et al., 2017) or
458 bacteriophage resistance (Martiny et al., 2015). We found that water temperature, reservoir depth,
459 and Nitrite concentration greatly affected the relative abundance of *Synechococcus* strains (Figure
460 S6, S7). However, we did not find significant linear or non-linear relationships between genetic
461 distance and preference for any of these abiotic niches. This was true when considering the whole
462 range of genetic distance (>93% identity) or a restricted range of more closely related strains
463 (>97%), suggesting that these traits are phylogenetically unconserved (at least within our dataset
464 and at the resolution of the 16S marker). Such a lack of conservation would be expected for a trait
465 such as Nitrite or Nitrate utilization in an organism capable of fixing nitrogen. Although some

466 marine *Synechococcus* can fix nitrogen (Zehr et al., 2001), further investigation will be needed to
467 determine if this is also the case in our freshwater community.

468 In contrast, total phosphorus (TP) preference was more phylogenetically constrained, even
469 though it explained less variation in community structure than Nitrite (Figure S6, S7).
470 Phosphorus use may therefore evolve "clock-like" along the *Synechococcus* phylogeny, following
471 an approximately linear model in the range of 96-100% nucleotide identity (Figure 1). This
472 contrasts with the observation in another cyanobacterium (*Prochlorococcus*) that organic
473 phosphate niches are fast-evolving and thus correlated with the phylogeny only near the tips of
474 the phylogenetic tree (Coleman & Chisholm 2010; Martiny et al., 2015). Consistent with this
475 idea, we identified significant niche separation among very closely related (~99% identical in
476 16S) *Synechococcus*, although the effect was driven by only three pairs of strains (Figure S12).
477 We also note that TP is a fairly crude measurement, encompassing both organic and inorganic
478 forms of phosphorus, and adaptation to these finer-grained niche dimensions may evolve
479 differently in *Synechoccus.* TP may not even be the driving selective pressure and may simply be
480 a marker of another important factor (*e.g.* runoff, overall biomass, etc.). Furnas Reservoir is
481 meso-eutrophic and phosphorus is probably not limiting (concentration varies from 8 to 53µg/L,
482 with a mean ~25µg/L). This might explain why only 77 of the 780 *Synechococcus* strain pairs
483 showed a significant co-response to phosphorus (Figure 1), indicating that the effect is driven by
484 a small number of closely-related strains with similar co-responses, and cannot be generalized to
485 all *Synechococcus*. This result might also be explained if *Synechococcus* represents a
486 polyphyletic group (Coutinho et al., 2016), composed of different clades with different
487 phosphorus niches. However, in our sample, *Synechococcus* is monophyletic (Figure S3, S4),
488 excluding this explanation. Overall, our results demonstrate how abiotic factors can shape
489 community composition within a genus but measured abiotic niches generally do not evolve

490 along the genus-level phylogeny.

491

492 Habitat filtering was originally defined to describe how abiotic factors select for genetically
493 similar (more closely related) organisms (Cadotte and Tucker 2017). Yet in practice, it is difficult
494 to disentangle the contributions of biotic and abiotic factors, and thus to satisfactorily define
495 habitat filtering (Kraft et al., 2015). In this study, we therefore estimated the influence of both
496 unmeasured abiotic variables and biotic factors, and found a negative association with genetic
497 distance (Figure 2). This suggests that more closely related *Synechococcus* tend to share similar
498 realized niches. As a result, we confirmed that genetically similar *Synechococcus* nodes tend to
499 co-occur, (Figure 3), as observed in studies using genomic (rather than 16S) similarity (Kamneva,
500 2017). This result could be unintuitive especially if *Synechococcus* nodes co-occur, but do not
501 share abiotic dimensions. For example, two taxa could be present in the same samples, but at very
502 different relative abundances. Therefore, they would co-occur but respond quantitatively to
503 different environmental variables. Overall, this result is consistent with habitat filtering, broadly
504 defined to include both biotic and abiotic factors.

505

506 Here we found that commonly-measured abiotic environmental factors (*e.g.* nutrients,
507 temperature) may be insufficient to capture relevant niche dimensions, and biotic co-occurrence
508 networks could be more informative about the evolution of niche preferences. In this study, we
509 measured abiotic factors such as temperature and nutrients, which are important in structuring
510 microbial communities, and yet these factors are largely insufficient to account for realized
511 niches or their distribution across the *Synechococcus* phylogeny. However, more closely-related
512 *Synechococcus* tend to co-occur with more similar surrounding microbial communities. As
513 previously suggested, the surrounding microbial community could provide hints as to the nature

514 of a realized niche (Cohan and Koeppel, 2008; Tromas et al., 2018). For example, the presence of
515 certain taxa could indicate the importance of an unmeasured abiotic factor (*e.g* Acidophiles or
516 Alkaliphiles as indicators of pH). In some cases, co-occurrence may indicate real biological
517 interactions (*e.g.* physical association, predation, cross-feeding, etc.). Some of the bacteria whose
518 co-occurrence with *Synechococcus* was strongly phylogenetically conserved have previously
519 been found in physical association with algae and cyanobacteria (Table S1; Mori et al. 2015;
520 Fullbright et al., 2018) and future work could experimentally test the nature of possible
521 interactions with *Synechococcus* strains. Alternatively, co-occurring microbes could simply share
522 similar unmeasured niches (including unmeasured abiotic or biotic factors such as phage or
523 protozoan grazers), and mining their genome sequences or metabolic capabilities could provide
524 insight into the nature of these niches. In this case, considering biotic factors allowed us to
525 uncover a signal of habitat filtering within *Synechococcus* that would have been obscured by only
526 considering commonly measured abiotic factors.
527

References

Allewalt, J.P., Bateson, M.M., Revsbech, N.P., Slack, K., and Ward, D.M. (2006) Effect of Temperature and Light on Growth of and Photosynthesis by Synechococcus Isolates Typical of Those Predominating in the Octopus Spring Microbial Mat Community of Yellowstone National Park. *Appl Environ Microbiol* **72**: 544–550.

Becker, S., Richl, P., and Ernst, A. (2007) Seasonal and habitat-related distribution pattern of Synechococcus genotypes in Lake Constance. *FEMS Microbiol Ecol* **62**: 64–77.

Becraft, E.D., Cohan, F.M., Kühl, M., Jensen, S.I., and 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.

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** : 365.

Breiman L, Friedman JH, Olshen RA, Stone CJ. (1984) Classification and Regression Trees. Wadsworth International Group, Belmont, CA, USA.

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* **108**: 14288–14293.

Cade, B.S. and Noon, B.R. (2003) A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment* **1**: 412–420.

Cadotte, M.W. and Tucker, C.M. (2017) Should Environmental Filtering be Abandoned? *Trends in Ecology & Evolution* **32**: 429–437.

Callieri, C. (1996) Extinction coefficient of red, green and blue light and its influence on picocyanobacterial types in lakes at different trophic levels. *Memorie dell'Istituto Italiano di Idrobiologia* 54: 135-142.

Callieri, C., Caravati, E., Corno, G., and Bertoni, R. (2012) Picocyanobacterial community structure and space-time dynamics in the subalpine Lake Maggiore (N. Italy). *Journal of Limnology* **71**: e9–e9.

Callieri, C., Lami, A., and Bertoni, R. (2011) Microcolony Formation by Single-Cell Synechococcus Strains as a Fast Response to UV Radiation. *Appl Environ Microbiol* **77**: 7533–7540.

Cohan, F.M. and Koeppel, A.F. (2008) The origins of ecological diversity in prokaryotes. *Curr Biol* **18**: R1024-1034.

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.

Coutinho, F., Tschoeke, D.A., Thompson, F., and Thompson, C. (2016) Comparative genomics of Synechococcus and proposal of the new genus Parasynechococcus. *PeerJ* **4**: e1522.

De'ath G. (2007) mvpart: Multivariate partitioning. R package version 1.6-2.

Dhariwal, A., Chong, J., Habib, S., King, I.L., Agellon, L.B., and Xia, J. (2017) MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* **45**: W180–W188.

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

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9**: 968–979.

Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. (2012) Microbial Co-occurrence Relationships in the Human Microbiome. *PLOS Computational Biology* **8**: e1002606.

Fernandes, A.D., Reid, J.N., Macklaim, J.M., McMurrough, T.A., Edgell, D.R., and Gloor, G.B. (2014) Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**: 15.

Ferris, M.J., Kühl, M., Wieland, A., and Ward, D.M. (2003) Cyanobacterial Ecotypes in Different Optical Microenvironments of a 68°C Hot Spring Mat Community Revealed by 16S-23S rRNA Internal Transcribed Spacer Region Variation. *Appl Environ Microbiol* **69**: 2893–2898.

Fierer, N. and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.

Friedman, J. and Alm, E.J. (2012) Inferring Correlation Networks from Genomic Survey Data. *PLOS Computational Biology* **8**: e1002687.

Fulbright, S.P., Robbins-Pianka, A., Berg-Lyons, D., Knight, R., Reardon, K.F., Chisholm, S.T. (2018) Bacterial community changes in an industrial algae production system. *Algal Res* **31**:147–156.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.

621 Green, J.L., Bohannan, B.J.M., and Whitaker, R.J. (2008) Microbial Biogeography: From
622 Taxonomy to Traits. *Science* **320**: 1039–1043.
623
624 Guedes, I.A., Rachid, C.T.C.C., Rangel, L.M., Silva, L.H.S., Bisch, P.M., Azevedo, S.M.F.O.,
625 and Pacheco, A.B.F. (2018) Close Link Between Harmful Cyanobacterial Dominance and
626 Associated Bacterioplankton in a Tropical Eutrophic Reservoir. *Front Microbiol* **9**: 424.
627
628 Horner-Devine, M.C. and Bohannan, B.J.M. (2006) Phylogenetic Clustering and Overdispersion
629 in Bacterial Communities. *Ecology* **87**: S100–S108.
630
631 Hui, F.K.C., Taskinen, S., Pledger, S., Foster, S.D., and Warton, D.I. (2015) Model-based
632 approaches to unconstrained ordination. *Methods in Ecology and Evolution* **6**: 399–411.
633
634 Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harb Symp Quant Biol* **22**: 415–427.
635
636 Hutchinson, G.E. (1961) The Paradox of the Plankton. *The American Naturalist* **95**: 137–145.
637
638 Kamneva, O.K. (2017) Genome composition and phylogeny of microbes predict their co-
639 occurrence in the environment. *PLOS Computational Biology* **13**: e1005366.
640
641 Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7:
642 improvements in performance and usability. Mol Biol Evol **30**: 772–780.
643
644 Koeppel, A.F. and Wu, M. (2014) Species matter: the role of competition in the assembly of
645 congeneric bacteria. *ISME J* **8**: 531–540.
646
647 Koenker, R. (2015). Package 'quantreg'
648
649 Koenker, R., Machado, J.A.F. (1999). Goodness of fit and related inference processes for quantile
650 regression. Journal of the American Statistical Association. **94**, 1296-1310.
651
652 Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., and Levine, J. M. (2015).
653 Community assembly, coexistence and the environmental filtering metaphor. Funct. Ecol.
654 **29**, 592–599.
655
656 Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics
657 Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870–1874.
658
659 Kurmayer, R., Christiansen, G., and Chorus, I. (2003) The Abundance of Microcystin-Producing
660 Genotypes Correlates Positively with Colony Size in Microcystis sp. and Determines Its
661 Microcystin Net Production in Lake Wannsee. *Appl Environ Microbiol* **69**: 787–795.
662
663 Larkin, A.A and Martiny, A.C. (2017) Microdiversity shapes the traits, niche space, and
664 biogeography of microbial taxa. Environmental Microbiology Reports **9**:55-70.
665
666 Letten, A.D., Keith, D.A., Tozer, M.G., and Hui, F.K.C. (2015) Fine-scale hydrological niche
667 differentiation through the lens of multi-species co-occurrence models. *Journal of*

Ecology ***103***: 1264–1275.

Magurran, A.E. and Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**: 714.

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.

Mori, J.F., Neu, T.R., Lu, S., Händel, M., Totsche, K.U., and Küsel, K. (2015) Iron encrustations on filamentous algae colonized by Gallionella-related bacteria in a metal-polluted freshwater stream. *Biogeosciences* **12**: 5277–5289.

Needham, D. M., Fuhrman, J. A. (2016) Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature Microbiology*, **1**: 16005.

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.

Nusch, E.A. (1980) Comparison of different methods for chlorophyll and phaeopigment determination. Arch. Hydrobiol. Beih. **14** : 14–36."

Pascual-García, A., Tamames, J., and Bastolla, U. (2014) Bacteria dialog with Santa Rosalia: Are aggregations of cosmopolitan bacteria mainly explained by habitat filtering or by ecological interactions? *BMC Microbiol* **14**: 284.

Preheim, S.P., Perrotta, A.R., Martin-Platero, A.M., Gupta, A., and Alm, E.J. (2013) Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl Environ Microbiol* **79**: 6593–6603.

Price MN, Dehal PS, Arkin AP. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**:1641–1650.

Rohwer, R.R., Hamilton, J.J., Newton, R.J., and McMahon, K.D. (2018) TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution. *mSphere* **3**: e00327-18.

Röttjers, L. and Faust, K. (2018) From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol Rev* **42**: 761–780.

Scanlan, D.J. (2003) Physiological diversity and niche adaptation in marine Synechococcus. In, *Advances in Microbial Physiology*. Academic Press, pp. 1–64.

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.

Steunou, A.-S., Bhaya, D., Bateson, M.M., Melendrez, M.C., Ward, D.M., Brecht, E., et al. (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. Proc Natl Acad Sci USA **103**: 2398–2403.

Stockner, J., Callieri, C., and Cronberg, G. (2002) Picoplankton and Other Non-Bloom-Forming Cyanobacteria in Lakes. In, Whitton,B.A. and Potts,M. (eds), *The Ecology of Cyanobacteria: Their Diversity in Time and Space*. Dordrecht: Springer Netherlands, pp. 195–231.

Tromas, N., Taranu, Z.E., Martin, B.D., Willis, A., Fortin, N., Greer, C.W., and Shapiro, B.J. (2018) Niche Separation Increases With Genetic Distance Among Bloom-Forming Cyanobacteria. *Front Microbiol* **9** : 438

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. The *ISME Journal* **11**: 1746–1763.

Vörös, L., Callieri, C., Balogh, K.V., and Bertoni, R. (1998) Freshwater picocyanobacteria along a trophic gradient and light quality range. *Hydrobiologia* **369**: 117–125.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., and Hui, F.K.C. (2015) So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution* **30**: 766–779.

Webb, C.O., Ackerly, D.D., McPeek, M.A., and Donoghue, M.J. (2002) Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics* **33**: 475–505.

Zheng, Q., Wang, Y., Xie, R., Lang, A.S., Liu, Y., Lu, J., et al. (2018) Dynamics of Heterotrophic Bacterial Assemblages within Synechococcus Cultures. *Appl Environ Microbiol* **84**: e01517-17.

Data availability

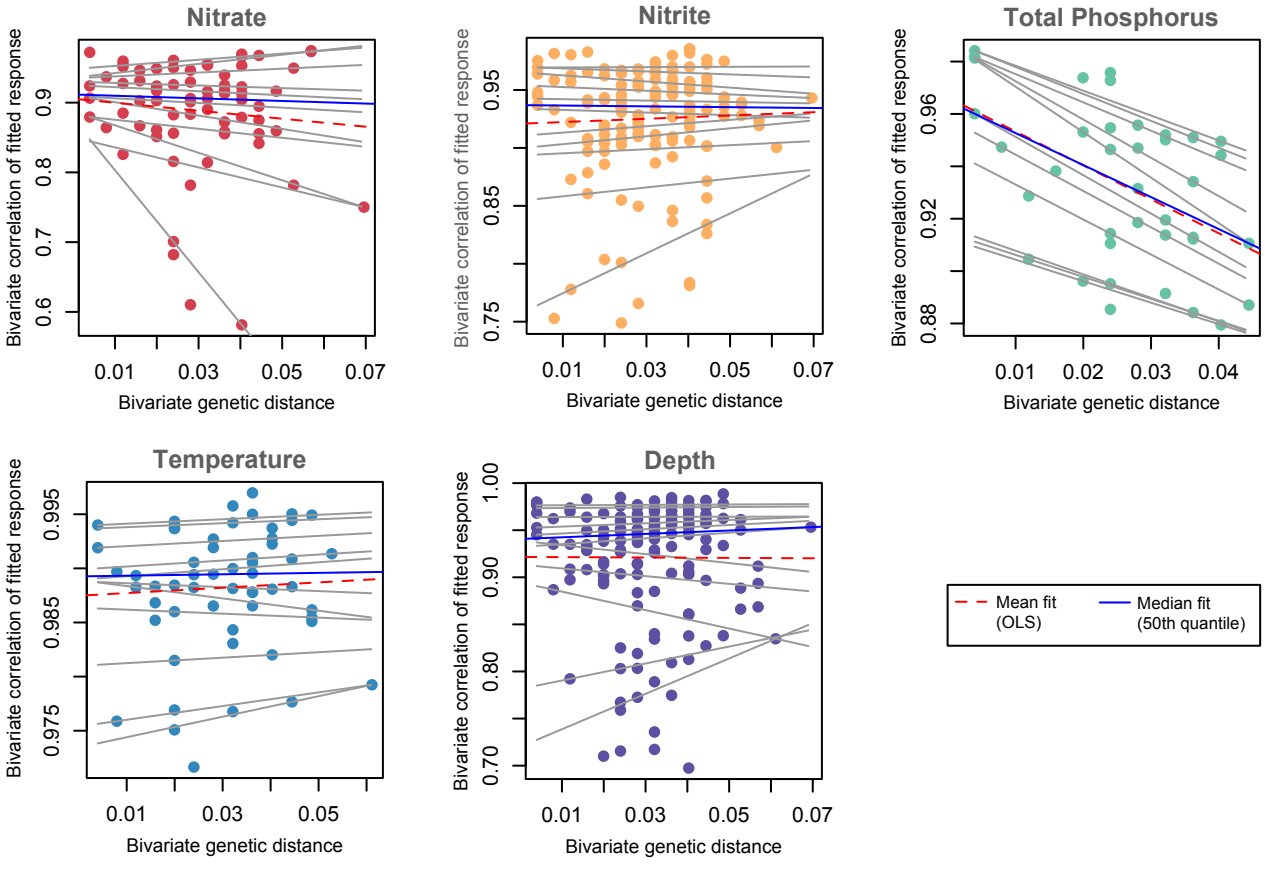
Raw sequence data have been deposited NCBI GenBank under BioProject number PRJNA544938.

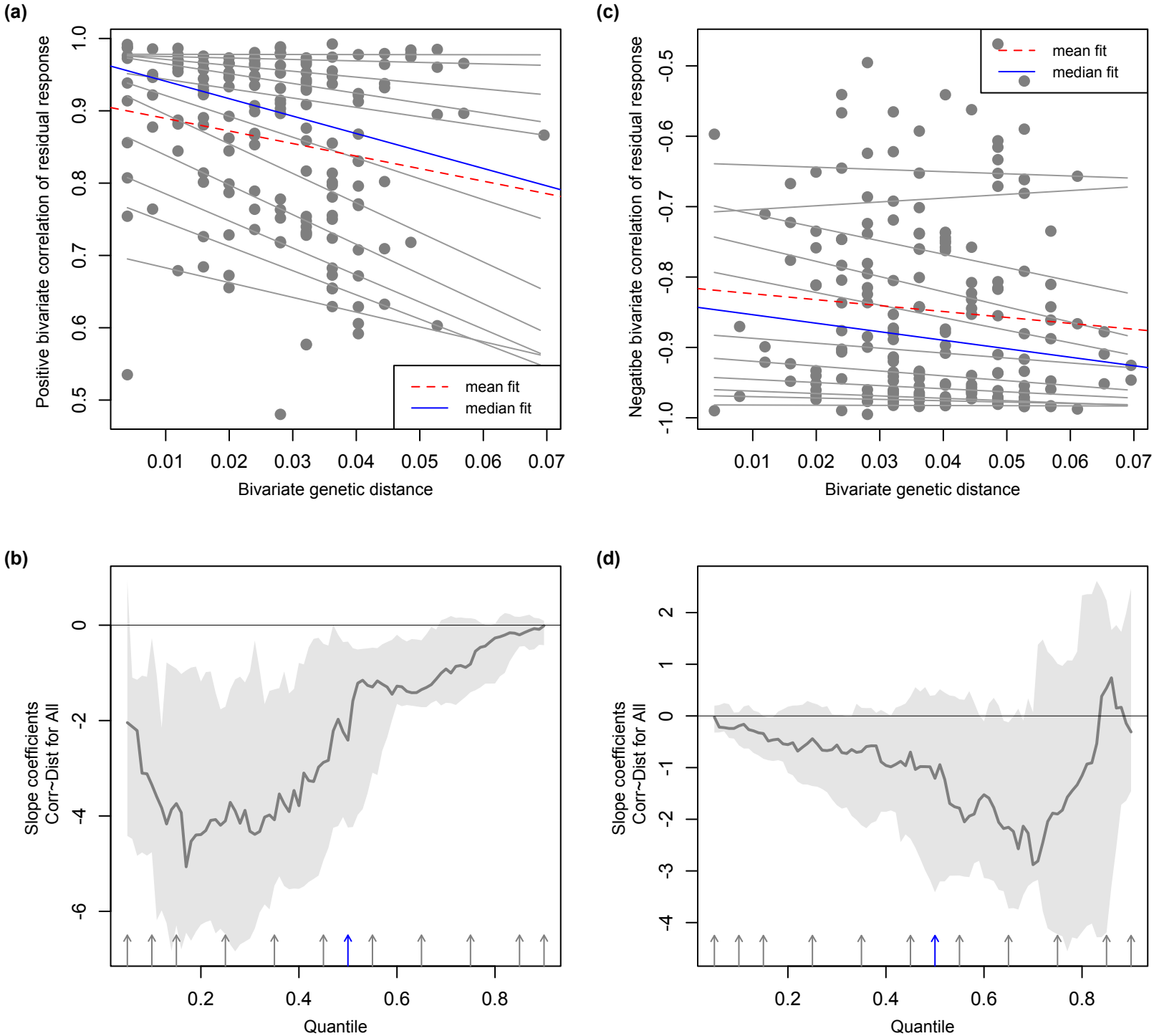
Conflict of Interest

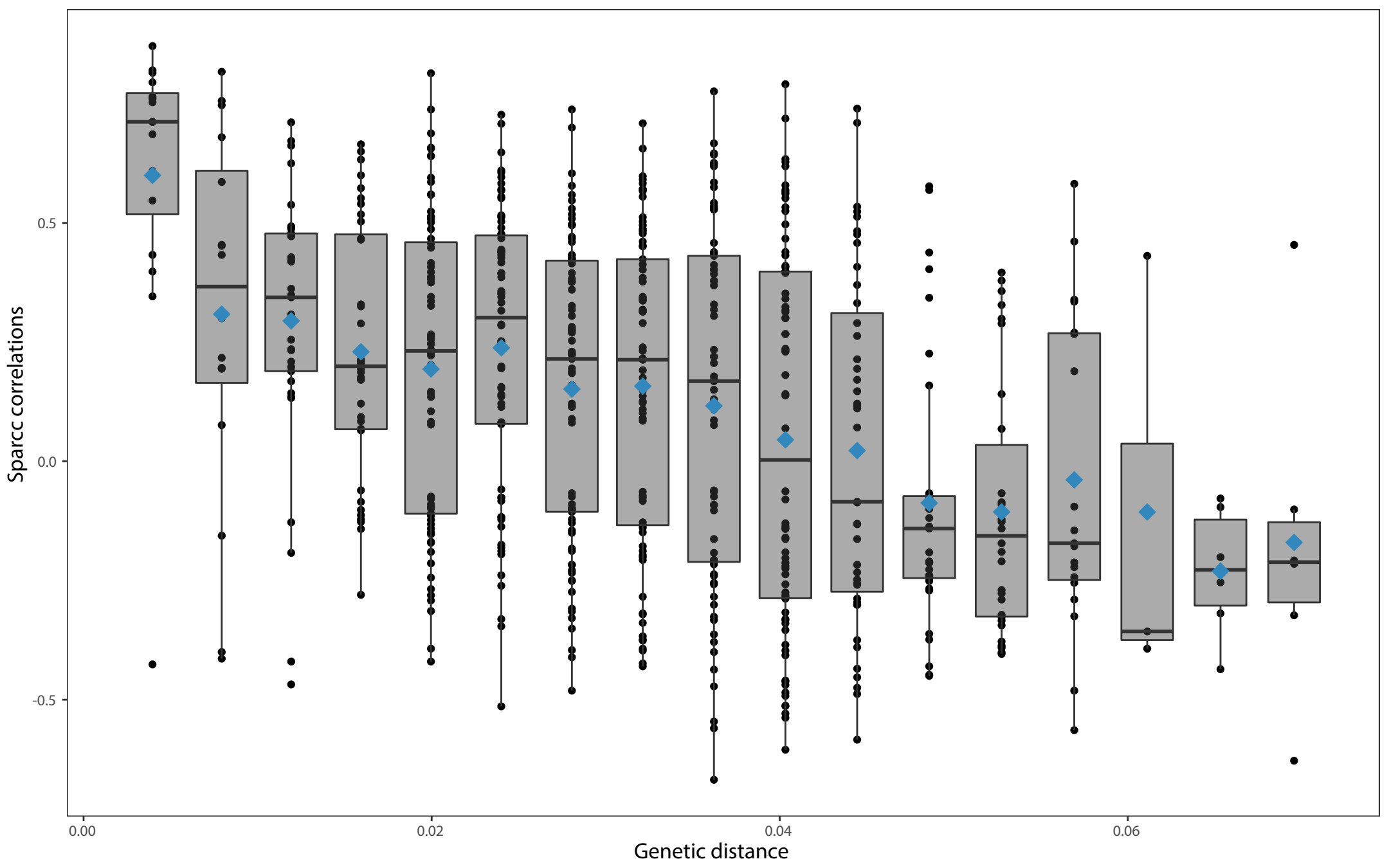
The authors declare no conflict of interest.

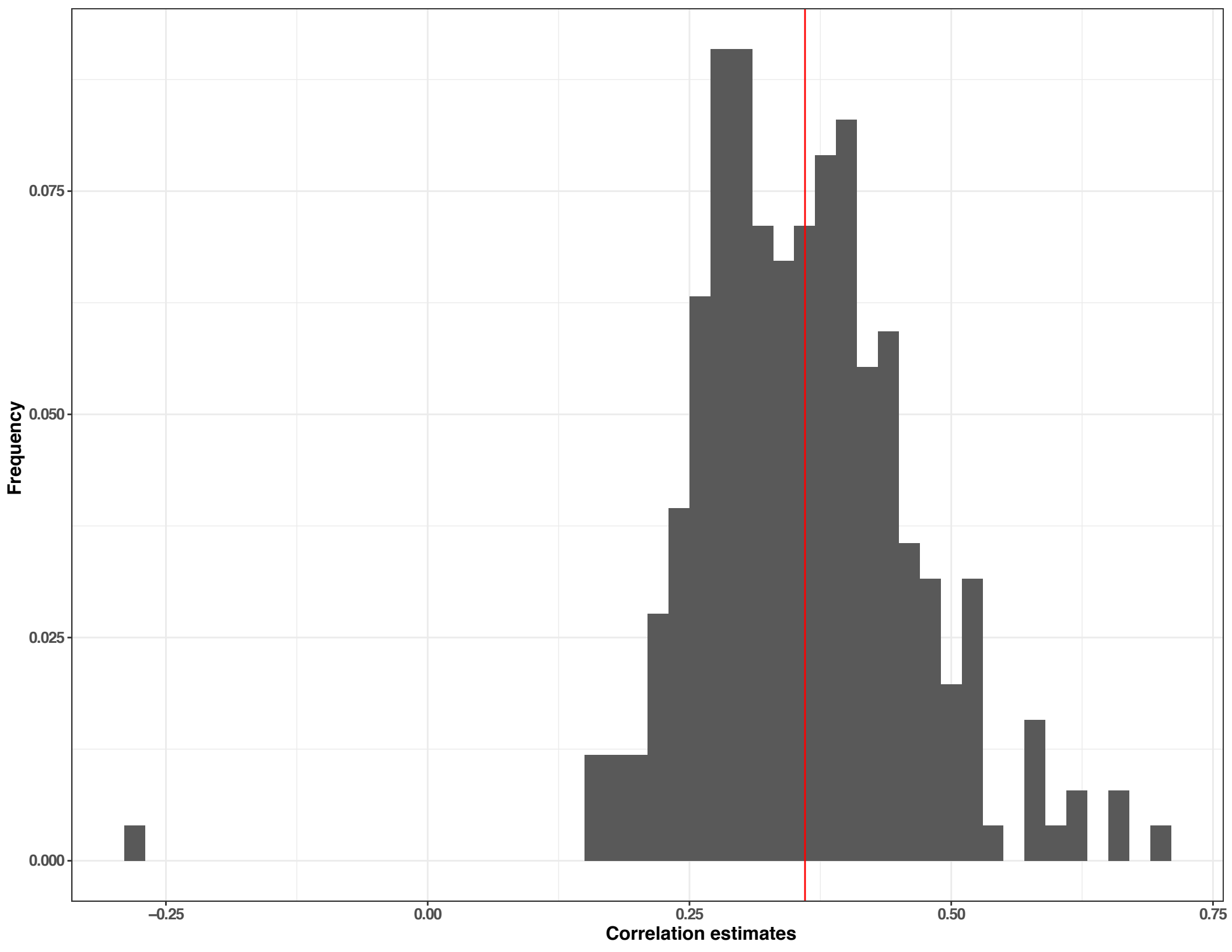
Acknowledgments

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. BJS was supported by a Canada Research Chair and NSERC Discovery Grant. Sampling and samples processing were supported by a grant from Furnas Centrais Hidroelétrica S.A. (Brazil) to AG. JSMP and DAP received a scholarship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). A sabbatical visit of AG to the Université de Montréal AG was supported by a CAPES Senior Fellowship.


Nitrate
Nitrite
Total Phosphorus
Temperature
Depth
Bivariate correlation of fitted response
Bivariate genetic distance
Mean fit (OLS)
Median fit (50th quantile)


(a)
Positive bivariate correlation of residual response
1.0
0.9
0.8
0.7
0.6
0.5
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Bivariate genetic distance
mean fit
median fit
(c)
Negatibe bivariate correlation of residual response
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Bivariate genetic distance
mean fit
median fit
(b)
Slope coefficients
Corr~Dist for All
0
-2
-4
-6
0.2
0.4
0.6
0.8
Quantile
(d)
Slope coefficients
Corr~Dist for All
2
1
0
-1
-2
-3
-4
0.2
0.4
0.6
0.8
Quantile


Sparcc correlations
0.5
0.0
-0.5
0.00
0.02
0.04
0.06
Genetic distance


Frequency
0.075
0.050
0.025
0.000
−0.25
0.00
0.25
0.50
0.75
Correlation estimates