The evolution of biotic and abiotic realized niches within freshwater *Synechococcus*

Authors

Nicolas Tromas[1#], Mathieu Castelli[2#], Zofia E. Taranu[3], Juliana S. M. Pimentel[4], Daniel A. Pereira[4], Romane Marcoz[5], Alessandra Giani[4] and B. Jesse Shapiro[1*]

SUPPLEMENTARY TABLE

Nodes	Correlation estimate*	P-value	Taxonomy
2344	**0.653**	**1.67e-4**	k_Bacteria;p_Planctomycetes;c_Planctomycetia;o_Gemmatales; f_Gemmataceae;g_Gemmata;s__unclassified
4557	**0.651**	**1.06e-10**	k_Bacteria ;p_Verrucomicrobia;c_[Pedosphaerae];o_[Pedosphaerales];f_auto67_4W;g_unclassified;s__unclassified
5008	**0.707**	**2.59e-5**	k_Bacteria;p_Proteobacteria;c_Deltaproteobacteria;o_Spirobacillales;f_unclassified;g_unclassified;s__unclassified
6401	**0.624**	**2.50e-3**	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Gallionellales; f_Gallionellaceae;g_Gallionella;s__unclassified
7445	**0.626**	**4.42e-5**	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Comamonadaceae;g_unclassified;s__unclassified
1169	**-0.283**	**0.0212**	k_Bacteria;p_Actinobacteria;c_Acidimicrobiia;o_Acidimicrobiales; f_C111;g_unclassified;s__unclassified
*Correlation between |_r| and genetic distance			

Table S1. Non-*Synechococcus* taxa with highest positive and negative correlation estimates.

SUPPLEMENTARY FIGURES

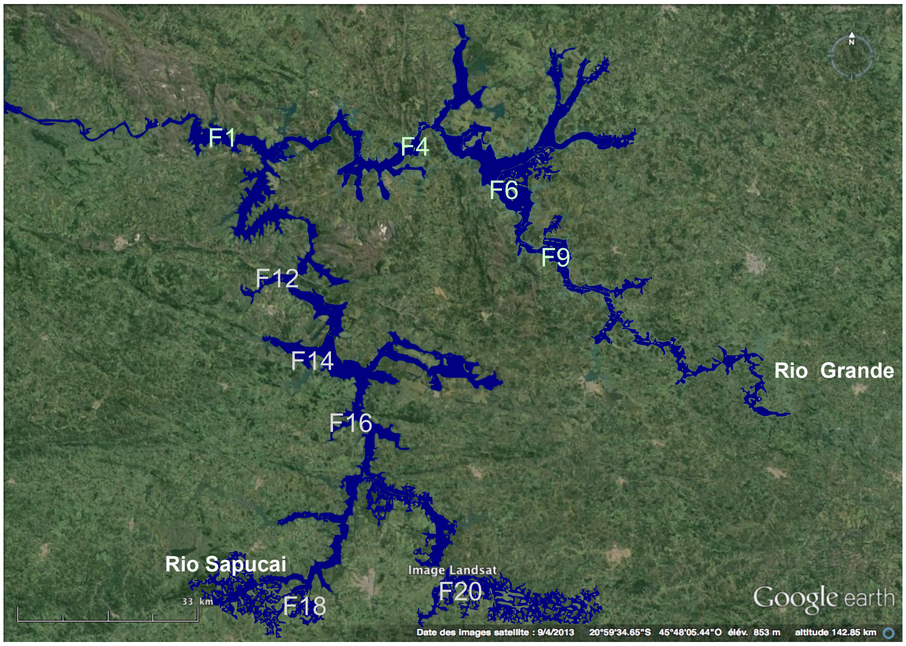

Figure S1. Map of the sampling stations. (Google Earth). Furnas reservoir is located in South-Eastern Brazil (20o40’S; 46o19’W). This large reservoir (1,440 Km^2; 20.95 billion m^3) is composed of two major rivers, Grande and Sapucaí, and several small tributaries. F12-F20 stations are mainly located in regions occupied by livestock, agricultural activities and several small to medium cities. F4 –F9 are located in less impacted areas. F1 is located close to the dam, in the intersection of both branches.

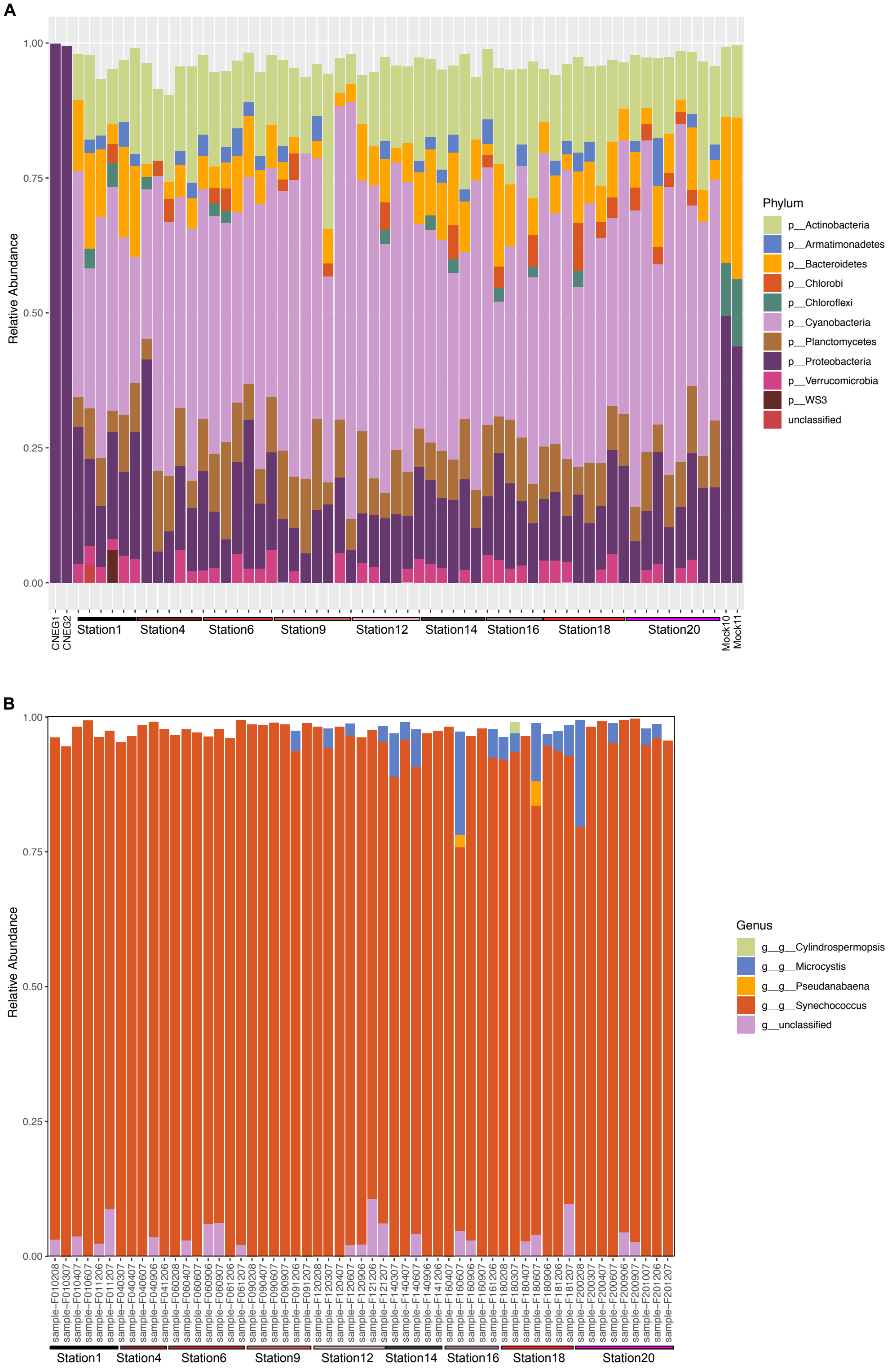

Figure S2. Barplot of the relative abundance of microbial taxa across samples in Furnas reservoir.

A. Barplot of the microbial composition at the phylum level, including negative controls (left) and mock communities (right). **B.** Barplot of the cyanobacterial composition at the genus level.

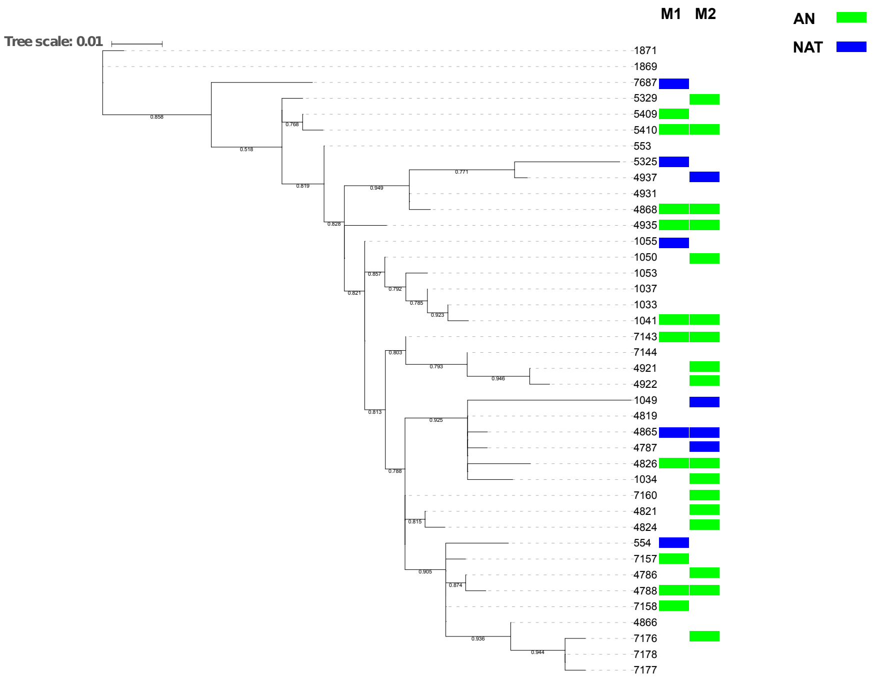

Figure S3. Phylogenetic distribution of *Synechococcus* nodes across less impacted (NAT, F1-F9) and more impacted (AN, F12-20) branches of the reservoir. Two different methods (ALDEx2; M1 and DESeq2; M2) were used to determine if a *Synechococcus* node is differentially associated with one of these two branches.

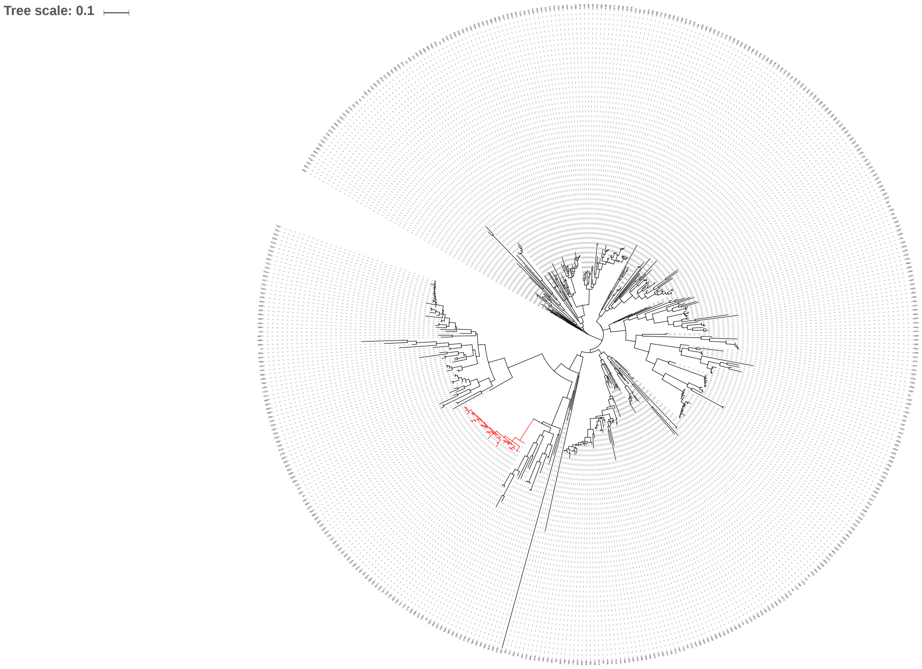

Figure S4. Phylogenetic characterization of 16S rRNA gene from the Furnas reservoir. *Synechococcus* nodes are highlighted in red, forming a monophyletic group.

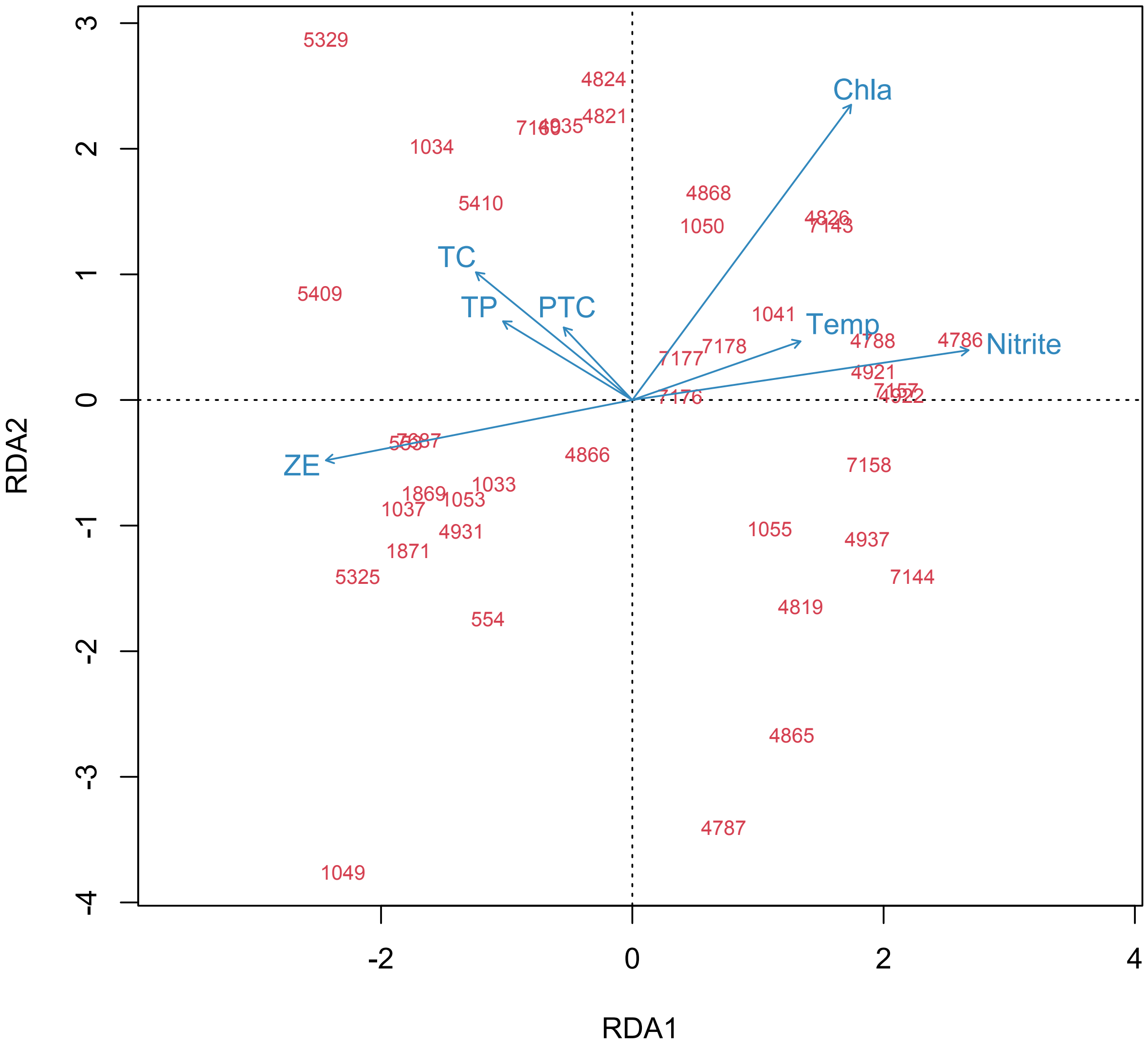

Figure S5. Parsimonous redundancy analysis of the *Synechococcus* community versus the forward-selected environmental variables. Forward selection selected Chl *a*, nitrite, total phosphosrus (TP), depth (ZE), TC, PTC and water temperature (Temp) as important variables, whereby the first two constrained axes (RDA 1 and RDA 2) explained 40% of the variance observed in the *Synechococcus* population (*Synechococcus* nodes are represented in red).

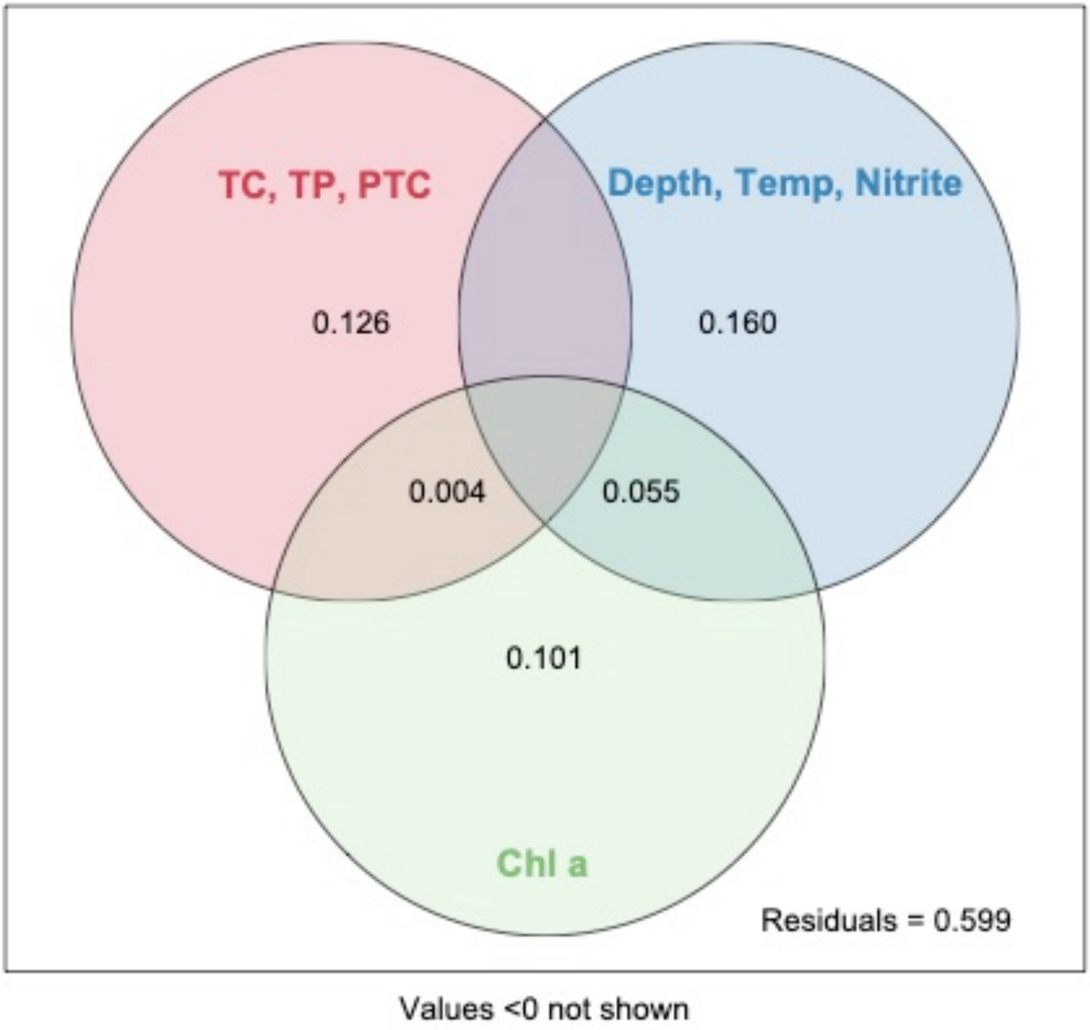

Figure S6. Variance partitioning analysis of the forward-selected environmental variables. Based on the RDA (Fig. S1) we lumped groups of explanatory variables that best separated the first two RDA dimensions (i.e., group 1 = TC, TP, PTC ; group 2 = Depth, Temp, Nitrite ; group 3 = Chl a) and tested the proportion of the *Synechococcus* population that was jointly and uniquely explained by each group of variables. Depth, temperature and nitrite alone explained the greatest, unique amount (16%) of the *Synechococcus* community.

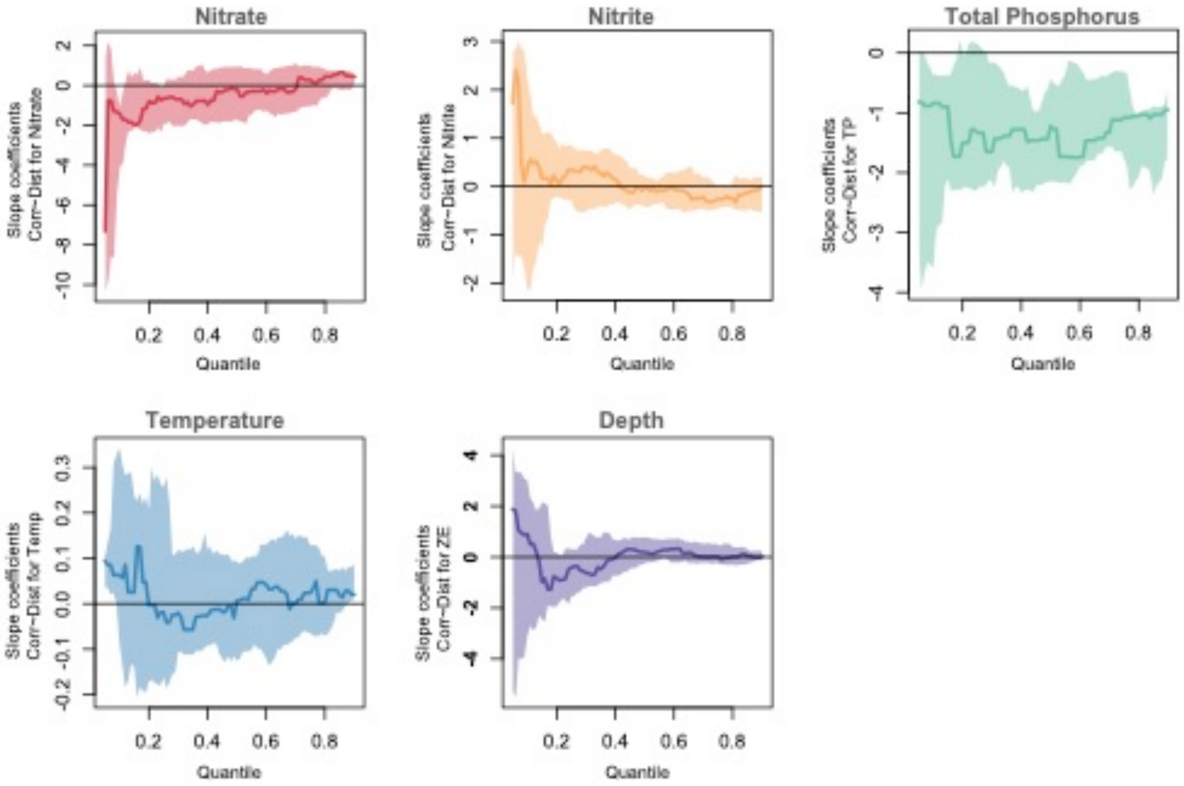

Figure S7. Significance plot of the quantile regression models of the degree of co-response vs genetic distance. Quantile slope estimates (dark grey line) and corresponding confidence intervals (light grey bands) of the LVM co-responses vs genetic distance models across all quantiles of genetic distance are shown, where slopes at the 5th, 10th, 15th, 25th, 35th, 45th, 55th, 65th, 75th, 85th and 90th quantiles shown in Figure 2 are indicated by grey arrows. Significant slopes are those where both the slope estimate (dark grey) and confidence band (light grey) do not overlap with zero.

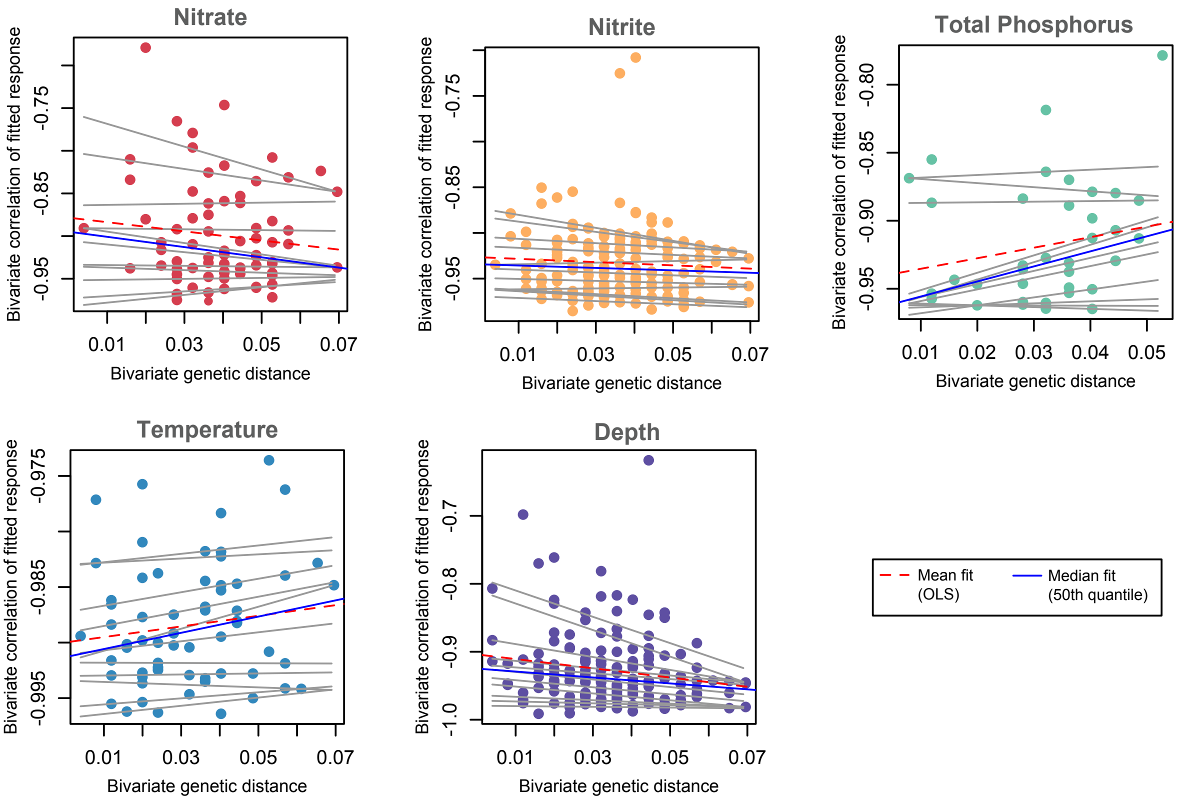

Figure S8. Scatterplot and quantile regression models of the degree of niche separation (negative correlation among LVM fitted values) vs. genetic distance. See Figure 1 for details.

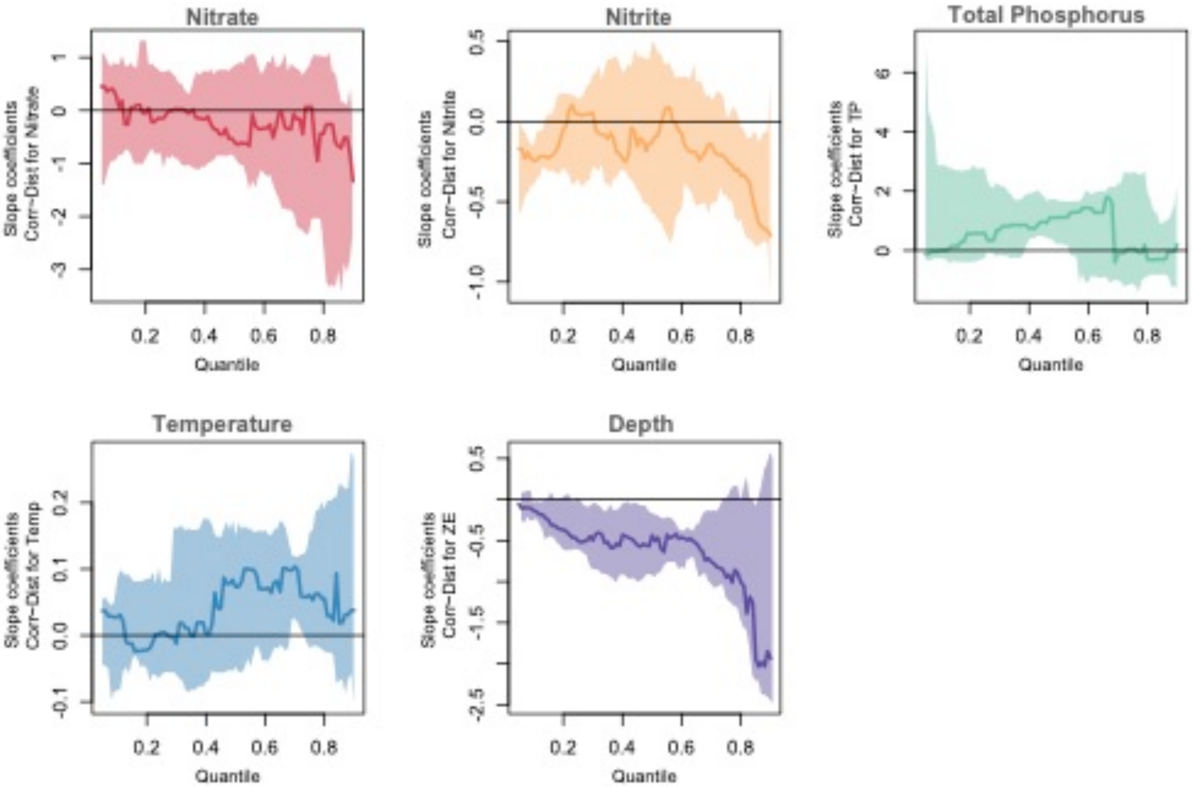

Figure S9. Significance plot of the quantile regression models of the degree of niche separation vs genetic distance. See Figure 2 for details.

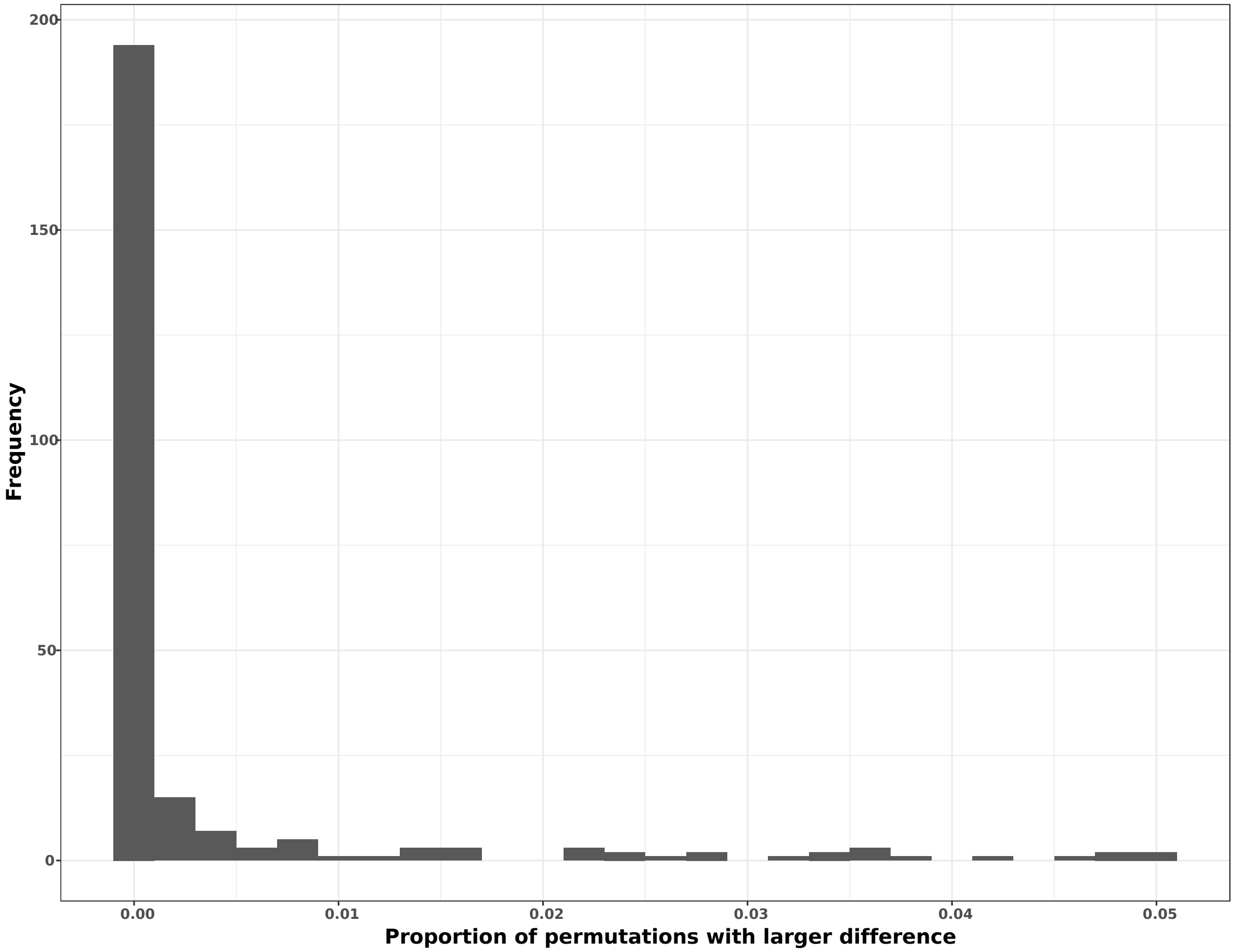

Figure S10. Permutation test of the association between *Synechoccus* genetic distance and association with similar surrounding communities. We estimated the proportion (p) of false positive correlations between genetic distance of *Synechococcus* nodes and the co-occurrence with non- *Synechococcus* taxa. The distribution is based on 1000 permutations for each of 253 non-*Synechococcus* taxa, randomizing the association between $|\Delta r|$ and genetic distance (Methods and R_script2). We then calculated p as the number of permutations yielding a larger correlation than observed (numerator) divided by the total number of permutations (denominator).