The global *Microcystis* interactome

Hooker, K.V.,[1,2] C. Li,[3], H. Cai[1], L.R. Krumholz,[3] K.D. Hambright,[1,2,*] H.W. Paerl,[4] M.M. Steffen,[5] A.E. Wilson,[6] M. Burford,[7] H.-P. Grossart,[8] D. P. Hamilton,[7,9] H. Jiang,[10] A. Sukenik,[11] D. Latour,[12] E.I. Meyer,[13] J. Padisák,[14] B. Qin,[10] R.M. Zamor,[15,a] and G. Zhu[10]

[1]Plankton Ecology and Limnology Laboratory and [2]Program in Ecology and Evolutionary Biology, Department of Biology, The University of Oklahoma, Norman, Oklahoma

[3]Department of Microbiology and Plant Biology and Institute for Energy and the Environment, The University of Oklahoma, Norman, Oklahoma

[4]Institute of Marine Sciences, The University of North Carolina at Chapel Hill, Morehead City, North Carolina

[5]Department of Biology, James Madison University, Harrisonburg, Virginia

[6]School of Fisheries, Aquaculture, and Aquatic Sciences, Auburn University, Auburn, Alabama

[7]Australian Rivers Institute, Griffith University, Nathan, Queensland, Australia

[8]Department of Experimental Limnology, Leibniz Institute for Freshwater Ecology and Inland Fisheries, Stechlin, and Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany

[9]Environmental Research Institute, University of Waikato, Waikato, New Zealand

[10]State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China

[11]Israel Oceanographic and Limnological Research, The Yigal Allon Kinneret Limnological Laboratory, Migdal, Israel

[12]LMGE, Laboratoire Microorganismes: Génome et Environnement, Clermont Université, Université Blaise Pascal, Aubière, France

[13] Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

[14]Department of Limnology, Institute of Environmental Science, University of Pannonia, Veszprém, Hungary

[15]Grand River Dam Authority, Vinita, Oklahoma

[a]Present address: Department of Natural Sciences, Northeastern State University, Tahlequah, Oklahoma

*Correspondence: K. David Hambright, Department of Biology, 730 Van Vleet Oval, 304 Sutton Hall, University of Oklahoma, Norman, OK 73019, Phone: 405-325-7435, E-mail: dhambright@ou.edu

Email addresses:

K.V. Hooker: khooker@ou.edu

C. Li: chuangli@ou.edu

H. Cai: haiyuan_cai@ou.edu

L.R. Krumholz: krumholz@ou.edu

K.D. Hambright: dhambright@ou.edu

H.W. Paerl: hans_paerl@unc.edu

M.M. Steffen: steffemm@jmu.edu

A.E. Wilson: aew0009@auburn.edu

M. Burford: m.burford@griffith.edu.au

H.-P. Grossart: hgrossart@igb-berlin.de

D. P. Hamilton: david.p.hamilton@griffith.edu.au

H. Jiang: hljiang@niglas.ac.cn

A. Sukenik: assaf@ocean.org.il

D. Latour: delphine.latour@univ-bpclermont.fr

E.I. Meyer: meyere@uni-muenster.de

J. Padisák: padisak@almos.uni-pannon.hu

B. Qin: qinbq@niglas.ac.cn

R.M. Zamor: rmzamor@gmail.com

G. Zhu: gwzhu@niglas.ac.cn

Running head: *Microcystis* interactome

Key words: bacteria, cyanobacteria, ecological biogeography, freshwater, HAB, microbiome, phycosphere, 16S rRNA taxonomy

Abstract

Bacteria play key roles in the function and diversity of aquatic systems, but aside from study of specific bloom systems, little is known about the diversity or biogeography of bacteria associated with harmful cyanobacterial blooms (cyanoHABs). CyanoHAB species are known to shape bacterial community composition and to rely on functions provided by the associated bacteria, leading to the hypothesized cyanoHAB interactome, a coevolved community of synergistic and interacting bacteria species, each necessary for the success of the others. Here, we surveyed the microbiome associated with *Microcystis aeruginosa* during blooms in twelve lakes spanning four continents as an initial test of the hypothesized *Microcystis* interactome. We predicted that microbiome composition and functional potential would be similar across blooms globally. Our results, as revealed by 16S rRNA sequence similarity, indicate that *M. aeruginosa* is cosmopolitan in lakes across a 280° longitudinal and 90° latitudinal gradient. The microbiome communities were represented by a wide range of OTUs and relative abundances. Highly abundant taxa were more related and shared across most sites and did not vary with geographic distance, thus, like *Microcystis*, revealing no evidence for dispersal limitation. High phylogenetic relatedness, both within and across lakes, indicates that microbiome bacteria with similar functional potential were associated with all blooms. While *Microcystis* and the microbiome bacteria shared many genes, whole-community metagenomic analysis revealed a suite of biochemical pathways that could be considered complementary. Our results demonstrate a high degree of similarity across global *Microcystis* blooms thereby providing initial support for the hypothesized *Microcystis* interactome.

Introduction

Seasonally-recurrent harmful algal blooms, particularly those of toxic cyanobacteria (cyanoHABs) are a global phenomenon of growing concern impacting water quality, ecosystem services, and human health associated with freshwater systems (Paerl and Otten 2013). Accelerating eutrophication and climate change (e.g., rising temperatures and shifting hydrological regimes) have resulted in the proliferation, intensification, and prolongation of cyanoHABs around the world (O'Neil et al. 2012; Carey et al. 2012; Paerl and Paul 2012; Mantzouki et al. 2018). Water quality in freshwater systems is intimately linked to anthropogenic activities. With rapidly expanding agricultural and urban development, as well as prolonged stratification periods due to global warming, many systems have become eutrophic or at risk of eutrophication. Although there is debate regarding the roles of specific nutrients in cyanoHAB dynamics (Conley et al. 2009; Paerl et al. 2011, 2016; Schindler 2012; Schindler et al. 2016), there is general agreement that increased nutrient inputs lead to increases in cyanobacterial biomass. CyanoHABs can alter ecosystem function by causing anoxia, depleting dissolved nutrients, and shifting zooplankton communities, all which alter carbon flows (Paerl and Otten 2013). Such blooms can produce hundreds of secondary metabolites, including hepatotoxins, neurotoxins, and dermatotoxic irritants, all of which can pose serious health threats to humans, livestock, and wildlife (Carmichael 2001; Huisman et al. 2018).

CyanoHABs, once thought to be homogeneous populations, are now known to be accompanied by a diverse suite of heterotrophic bacteria (Eiler and Bertilsson 2004; Steffen et al. 2012; Xu et al. 2018), which may play an important role in cyanobacterial bloom health and duration. First, Bell and Mitchell (1972) defined this potential interactive relationship between cyanobacteria and heterotrophic bacteria as the "phycosphere" (Paerl and Kellar 1978, 1979;

Paerl and Millie 1996). It is well known that cyanobacteria generate abundant dissolved organic carbon resources to the benefit of nearby heterotrophs that can subsequently return benefits to the cyanobacteria, including removal of reactive oxygen species, CO_2 generation, and nutrient recycling (Dziallas and Grossart 2011, 2012; Steffen et al. 2012; Paerl and Otten 2013). Moreover, *Microcystis* has been shown to alter ambient environmental conditions by decreasing oxygen concentrations and light availability (Paerl and Otten 2016), as well as by altering CO_2 and pH levels (Havens 2008), which is likely to affect nearby bacteria. Indeed, studies have found that cyanobacterial bloom species strongly impact bacterial community composition (e.g., *Nodularia* – Salomon et al. 2003, *Microcystis* – Li et al. 2011; Steiner et al. 2017). Specifically, heterotrophic bacteria can form close-knit aggregates with *Microcystis*, and studies have shown greater similarity between attached Bacteria and Archaea communities than between free-living assemblages during *Microcystis* blooms (Cai et al. 2014; Yang et al. 2017; Batista et al. 2018; Xu et al. 2018).

One reason for this close association could be that, like other Bacteria and Archaea (Swan et al. 2013; Giovannoni et al. 2014), Cyanobacteria have small genomes compared with eukaryotes (Herdman et al. 1979; Humbert et al. 2013). While this may be beneficial for rapid reproduction and evolution, it is not necessarily conducive for cyanoHAB formation. Genome reduction can lead to loss of functions (Giovannoni et al. 2014), but can also confer a selective advantage if the organism can obtain the lost function through a public good as described by Morris et al. (2012) in the Black Queen Hypothesis. This hypothesis suggests that natural selection can act on "leaky" functions where a public good is produced and available to the whole community. Coupled with selection towards smaller genomes to reduce replication-related fitness costs, some members of the community can receive metabolic products as public goods,

which are useful metabolites or other necessary resources that are leaked into the cell-external environment. With such products available extracellularly, these "leaky" functions become dispensable and once lost confer a selective advantage to that organism (Pande and Kost 2017). Garcia et al. (2015) proposed that this coevolved community of synergistic and interacting bacteria species was a bacterial community microbiome or "interactome", analogous to the microbiome concept described for humans (Human Microbiome Project Consortium 2012), soils (Fierer 2017), and coral reefs (Bourne et al. 2013). We hypothesize that the bacteria associated with *Microcystis* may be providing functions that help sustain it during blooms.

Given the small size of bacterial genomes and presumed ubiquity of *Microcystis aeruginosa* (Kützing) Kützing, if there exists a mutualistic interactome, we would predict a microbiome of a certain species of associated bacteria or perhaps metabolic functions to be preserved across geographically distinct *M. aeruginosa* blooms. We would also expect communities to be more similar than predicted by traditional biogeographic theory, where community similarity is expected to decrease with increasing geographic distance (MacArthur 1984; Nekola and White 1999; Green and Bohannan 2006; Nemergut et al. 2013). As a first test of this prediction, we examined community composition and function of the *M. aeruginosa* bloom microbiomes from 12 lakes across four continents, addressing the specific prediction that if global *Microcystis* blooms were composed of the same taxon, the blooms would support and require a similar suite of bacterial-provided functions, thus leading to highly similar bacterial communities across *Microcystis* blooms regardless of geographic location.

Methods

Sample Collection

Samples were collected from 12 lakes spanning the globe during the peak of the *Microcystis* blooms between May 2016 and July 2017 (Fig. 1). Three surface water samples were taken from each lake with a clean 1000-mL beaker or Erlenmeyer flask and set aside undisturbed for 10 min to allow the cyanobacteria to float to the surface. Concentrated *Microcystis* biomass was poured off the top of the flask or beaker through a Nitex screen (100 µm pore size) stretched between a PVC pipe and PVC coupler to collect large *Microcystis* aggregates on the screen. Each Nitex screen was rolled up using forceps and transferred into 2-ml screw cap tubes containing 1.0 ml DNA preservative (DNA/RNA Shield, Zymo Research, Irvine, CA, United States). This process was carried out for each water sample. Tubes were stored at –20 °C until shipping. They were shipped at ambient temperature and once received were stored at –20 °C prior to extraction.

DNA extraction and sequencing

DNA extraction from the preserved samples was performed using Zymo Research *Quick*-DNA Fecal/Soil Microbe Miniprep Kits (Zymo Research, Irvine, CA, United States) following the manufacturer's recommended protocol. The procedure involved placing the Nitex screen into the lysis tube with glass beads, followed by mixing with a Bead Beater vortex mixer (BioSpec products, Bartlesville, OK) for 2 min as a first step in the extraction process. Extracted DNA was collected by decanting and used as a template for amplifying bacterial 16S rRNA gene through PCR. Forward (S-D-Bact-0341-b-S-17, 5'-CCTACGGGNGGCWGCAG-3') and reverse primers (S-D-Bact-0785-a-A- 21, 5'-GACTACHVGGGTATCTAATCC-3') were used for targeting the V3 and V4 regions of bacterial 16S rRNA gene (Klindworth et al. 2013). Each 50-µL PCR reaction mix contained 2 µL of template (~30 ng), 2 µL of each primer (0.4 uM, final concentration), and 25 µL PCR master mix containing DreamTaq polymerase (Thermo Fisher

Scientific, USA). Thermocycler conditions were as follows: 94 °C for 3 min followed by 28 cycles of 94 °C for 30 sec, 55 °C for 60 sec, and 72 °C for 75 sec, and a final elongation step at 72 °C for 10 min. PCR products were visualized on 1% agarose gels to confirm amplification and then purified by QIAquick PCR purification kit (Qiagen, Valencia, CA, United States) to remove primers. From each purified sample, 4 µL were added to a second PCR mixture containing barcoded primers for multiplexed Illumina sequencing. Through re-amplification for another eight cycles in the second PCR reaction, each sample received a unique "barcode" sequence as previously described (Wawrik et al. 2012). The secondary PCR products were quantified with the Qubit dsDNA BR Assay kit (Life Technologies, Carlsbad, CA, USA) on a Qubit 2.0 Fluorometer. Amplicons of all samples were pooled in an equimolar amount. Pooled samples were purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and re-quantified with the Qubit. Paired-end sequencing of the library was performed at the Oklahoma Medical Research Foundation, using the MiSeq Reagent Kit (v3) with the read length set to 2 × 300 base pairs (bp).

Shotgun metagenomics was used to profile functional potential and to recover whole genome sequences from the *Microcystis* and microbiome communities. Metagenomic sequencing was performed on two replicates per sample. A 300-bp paired-end library was constructed according to the instructions from Illumina. The libraries were sequenced on an Illumina Genome Analyzer IIx at the Oklahoma Medical Research Foundation. Eighteen metagenomes (two per lake) were multiplexed on two lanes, and a median total of ~48 million raw paired-end reads was obtained for each sample (range: ~32–96 million, due to variations in library loading).

Sequence Processing and Analysis

The 16S raw sequence data was processed via the QIIME pipeline (V1.9.1) (Caporaso et al. 2010), integrated with UPARSE-OTU algorithm. Paired-end reads were joined using the join_paired_ends.py function according to the SeqPrep method (https://github.com/jstjohn/SeqPrep) with a minimum overlap of 150 bp. The quality-joined fragments were demultiplexed and primer sequences removed. After using the FASTQ Quality Filter (q = 20, p = 80) to remove unqualified sequences, the remaining fragments were clustered into operational taxonomic units (OTUs) at the 97% similarity level with UPARSE OTU clustering (Edgar 2013) which generates a representative set of high-quality OTU sequences and filters out chimeric sequences via the de novo mode (using usearch 11.0.667 Edgar 2010). Taxonomic annotations were assigned to each high-quality OTU sequence by RDP's naive Bayesian rRNA Classifier (Wang et al. 2007) against the SILVA SSUv132 Reference Database (Quast et al. 2012), at the confidence threshold of 80%. To perform phylogenetic analysis, OTU sequences were also aligned against the SILVA database with PyNAST (Caporaso et al. 2009), filtered, and a phylogenetic tree constructed with FastTree (Price et al. 2010). Operational taxonomic units that failed in alignment or were classified as either eukaryote, archaea, chloroplast, or mitochondria were discarded from the OTU table, as were OTUs with fewer than 100 counts summed across samples.

To compare the community structure of the *Microcystis* microbiome among lakes, we extracted non-cyanobacterial OTUs from the OTU table, and then retrieved their associated representative sequences. A phylogenetic tree with non-cyanobacterial OTU sequences was constructed, as detailed above. OTU tables without cyanobacteria were rarified to the number of

reads of the sample with the fewest reads. The rarified OTU table with pooled replicates was used for all downstream diversity calculation and statistical analysis.

Following quality trimming (Trimmomatic v 0.39; Bolger et al. 2014), with reads shorter than 20bp being discarded and removal of human genes (MetaWRAP; Uritskiy et al. 2018), the clean metagenomic data was assembled into contigs by de novo assembly of each sample sequence using metaSPAdes (SPAdes v.3.13.0; Bankevich et al. 2012). The interactome (i.e., *Microcystis* and its microbiome) metagenomic assembled genomes (MAGs) were generated for each lake using three tools with default options: MaxBin (v.2.2.6) (Wu et al. 2015), MetaBAT (v. 2.12.1) (Kang et al. 2015) and CONCOCT (v. 1.0.0) (Alneberg et al. - unpublished preprint doi: arXiv:1312.4038v1 [q-bio.GN]), followed by integration using DAS Tool (using a threshold of ≥ 70% genome completeness) (v. 1.1.1) (Sieber et al. 2018). The complete MAGs were then divided into two groups, *Microcystis* and heterotrophic bacteria, and pooled across the lakes. Protein-encoding genes of each group were annotated from the contigs with Prokka (v1.13.3) (Seemann 2014). Duplicate genes were removed with CD-HIT-EST (v 4.6) (Hahn et al. 2016). Genes were then converted into protein sequences using Prokka. The protein sequences of each group were annotated to Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologies to characterize individual gene functions using GhostKOALA (Kanehisa et al. 2016). To calculate the gene abundance in each sample, all KEGG annotated genes were first aligned with the clean reads by bowtie2 alignment software (Langmead and Salzberg 2012). The number of reads mapping to each gene was extracted using the SAMtools (v1.3.1) "idxstats" command. The abundance of each gene in all samples was calculated by get_count_table.py (https://github.com/edamame-course/Metagenome/blob/master/get_count_table.py). To compare the functions contributed by the microbiome to those of the *Microcystis* we constructed complete

KEGG pathways (no less than one missing gene). Due to the high phylogenetic and functional similarities found across lakes (see Results), we pooled *Microcystis* MAGs and microbiome MAGs (Li et al. 2018), respectively, in order to generate the pathways. We uploaded the KO numbers into the online KEGG mapper to construct the complete KEGG pathways.

Due to the low depth of coverage for *Microcystis* microbiome genes, we attempted to maximize the probability of identifying functional potential in the *Microcystis* microbiomes by repeating the process outlined above for MAGs using the total metagenomic data. We separated the *Microcystis* reads from the microbiome reads by aligning the clean metagenomic data with 16 *Microcystis* genomes (ten *M. aeruginosa*, one *M. viridis*, one *M. panniformis*, two *M. flos-aquae*, two *M. wesenbergii*) using bwa v0.7.15 (Li and Durbin 2009) and splitting the data based on this alignment. The reads were then assembled into contigs (SPAdes v3.1.1; Bankevich et al. 2012) and the contigs for each lake (*Microcystis* and the microbiome kept separate) were pooled across lakes into a single group and annotated (Prokka v1.13.3; Seemann 2014). Duplicate genes were then removed with CD-HIT-EST (v 4.6) (Hahn et al, 2009). The protein sequences of each group were annotated to KEGG orthologies, the gene abundance mapped to each lake, and the data pooled and complete KEGG pathways constructed as described above.

Statistical Analyses

All statistical analyses were completed in the R statistical environment v.3.5.1 (R Development Core Team 2018), except where noted otherwise. We first tested for associations between geographic distance and community dissimilarity across samples. We calculated the great circle geographic distance between sites using the "rdist.earth" function (fields v.9.6). We calculated Bray-Curtis dissimilarity using "vegdist" (vegan package v.2.5-3 (Oksanen et al.

2013). UniFrac values, weighted by abundance, were generated in QIIME. Generalized linear models (GLM) were used to assess the relationship between geographic distance and community dissimilarity measures (using default family = Gaussian, link = identity). Deviance explained by GLMs coupled with p-values was used to assess the significance and strength of the relationship.

To examine the phylogenetic relationship among *Microcystis* microbiome bacteria *within* each sample (α-diversity), we calculated mean-nearest-taxon-distance (MNTD) and the nearest-taxon-index (NTI) (Webb 2000) using the "mntd" and "ses.mntd" commands in the Picante package v.1.7 in R (Kembel et al. 2010; Swenson 2014). NTD is a measure of phylogenetic distance between each OTU within a sample and its closest relative in the same sample. The mean is then calculated across all phylogenetic distances in a sample to give a value of phylogenetic relatedness. To determine whether observed phylogenetic community composition was more or less related (or structured) than predicted by chance, null models were built by randomly shuffling the taxa within each community across the tips of the phylogeny (null.model = "taxa.labels" in "ses.mntd") and recalculating MNTD 999 times (Stegen et al. 2012). The resulting NTI (which is the negative output of "ses.mntd") distribution displays the number of standard deviations that the observed MNTD is from the mean of the null MNTD values. An αNTI value less than −2 indicates taxa within the community are more distantly related than by chance (phylogenetically overdispersed), a value greater than 2 indicates taxa are more closely related than expected by chance (phylogenetically clustered), and αNTI values between 2 standard deviations of 0 indicate that the observed community is no different from random (Webb 2000; Stegen et al. 2012). A two-sample t-test was used to assess whether the mean αNTI values from across all communities was significantly different from that of the null distribution.

We used beta-mean-nearest-taxon-distance (βMNTD) and beta-nearest-taxon-index (βNTI) to quantify the phylogenetic distance among samples (see Swenson 2014 for R code). Similar to its α-diversity analog, βMNTD is the mean of the phylogenetic distances between each OTU in a given sample to their nearest relative in the comparison sample. Null distributions were generated by randomizing OTUs across the phylogeny and recalculating βMNTD 999 times (Stegen et al. 2012). As before, βNTI is the number of standard deviations that the observed βMNTD is away from the mean of the null distribution. A value of βNTI less than −2 indicates more phylogenetic relatedness between samples than expected by chance, and a βNTI greater than 2 indicates less phylogenetic relatedness between the two communities than expected by chance (i.e., species in these communities are more phylogenetically distant from each other). A two-sample t-test was used to determine if the mean βNTI for two communities was significantly different from the null expectation.

The abundance of KEGG genes for the *Microcystis* microbiome for each lake (mapped from the pooled total metagenomic data) were used to measure the community functional dissimilarity using Bray-Curtis (pairwise comparisons between lakes). GLM were used to assess the relationship between geographic distance and community functional dissimilarity measures (using default family = Gaussian, link = identity). To explore important biogeochemical pathways that may be shared or different between the *Microcystis* and their associated bacteria, we compared the complete (no more than one gene missing) KEGG pathways between *Microcystis* and the microbiome. Due to the low depth of coverage of the microbiome, we used both the MAGs complete KEGG pathways and the total metagenome complete KEGG pathways for this analysis to maximize the probability of identifying a complete pathway.

317 ***Results***

318 The sampled lakes spanned a 280° longitudinal and 90° latitudinal gradient ranging from
319 444 to 11,777 km apart and represented *M. aeruginosa* blooms from four continents (Fig. 1).
320 After removal of Eukaryota, singletons, chloroplasts, and mitochondrial reads, 3.25 million high-
321 quality reads were represented by 454 unique (97% similarity) bacterial OTUs. *Microcystis* (927
322 total OTUs) was dominant in ten of the 12 lakes and ranged from 65-84 % of total sequence
323 abundance. Both Taihu and Wentowsee communities were made up of less than 50% *Microcystis*
324 (48% and 5%, respectively) sequences suggesting the lakes were not at peak bloom phase during
325 sampling. As such, these two lakes were removed from subsequent analyses. Castlerock was also
326 removed from subsequent analyses due to a loss of one of the triplicate samples. All *Microcystis*
327 OTUs with greater than 1000 total abundance were re-identified to species using NCIBI BLAST
328 function (Altschul et al. 1990) and were all identified as *M. aeruginosa*, accounting for 79-85%
329 of the total *Microcystis* across lakes and 53-68% of all OTU abundances across lakes. This
330 confirms that, at least at the level of 16S rRNA sequences, *M. aeruginosa* is a cosmopolitan
331 bloom-forming species.

332 The 454 non-cyanobacterial OTUs were associated with 35 bacterial classes, with most
333 OTUs identified as Alphaproteobacteria, Bacteroidia, Gammaproteobacteria, Clostridia,
334 Campylobacteria, Deltaproteobacteria, Negativicutes, Phycisphaerae, Gemmatimonadetes,
335 Acidobacteriia, Ignavibacteria, SM1A07, Melainabacteria, Cytophagia, Parcubacteria, and
336 Anaerolineae (Fig. 2b). The abundance and taxonomy of the associated bacteria differed among
337 sites (Fig. 2b) with Alphaproteobacteria, Bacteriodia, and Gammaproteobacteria classes
338 contributing most sequences across sites. Bray-Curtis dissimilarity values were relatively high
339 for all of the site comparisons indicating large differences in OTUs among sites (Fig. 3a).

Weighted UniFrac, on the other hand, was generally low, suggesting that among sites, there were many shared taxa (or a few shared taxa with high abundances) with some degree of phylogenetic relatedness (Fig. 3b). These results together suggest that there was a wide range of relative abundances of taxa across all of the sites, including many low abundance rare taxa (contributing to high Bray-Curtis values), while the highly abundant taxa were more related and shared across most sites (contributing to lower UniFrac values). Neither of the community dissimilarity metrics were significantly correlated with geographic distance (Fig. 3a and 3b; BC GLM deviance explained = 4.1%, $p = 0.2$; UF GLM DE = 0.97%, $p = 0.57$).

Phylogenetic α-diversity analysis using mean nearest taxon distance and null model generation revealed on average that within lakes, bacteria in the *Microcystis* microbiome were more phylogenetically related than predicted by chance (Fig. 4, + αNTI; $t = 10.13$, $p < 0.001$). Among lakes, the *Microcystis* microbiome communities also were significantly more phylogenetically related than expected by chance (Fig. 4, βNTI; $t = -12.65$, $p < 0.001$).

A total of 866 million metagenomic sequencing reads were generated from the nine lake *Microcystis*-microbiome communities. After trimming and filtering, 612 million (52.4 to 137.6 million per lake) clean reads were generated (Supplementary Table 1). Most of the clean metagenomic reads (55.2% to 76.6% per lake) belonged to *Microcystis*.

Analysis of only whole genome data revealed nine *Microcystis* genome bins (one MAG per lake) and but only 43 microbiome bacterial genome bins (3-10 MAGs per lake), indicating extremely low depth coverage for the microbiome bacteria. After removing the duplicate genes and RNA genes, 156,445 and 39,880 protein-coding genes were obtained from the microbiome and *Microcystis* genomes, respectively. Our analysis showed that 66,861 (~42%) of the microbiome genes, and 14,188 (~35%) of *Microcystis* genes were successfully assigned to the

363 KEGG orthology. By contrast, after removing duplicate and rRNA genes, analysis of the total
364 metagenomic data produced 407,658 and 54,312 protein-coding genes for the microbiome and
365 *Microcystis*, respectively, of which 139,384 (more than twice the number compared with MAGs)
366 and 17,817 (3,629 more), respectively, were successfully mapped to the KEGG orthology.
367 Microbiome gene diversity was similar across lakes (Fig. 5; low Bray-Curtis values) and did not
368 differ with geographic distance between the lakes (Fig. 5; BC GLM DE = 1.97%, p = 0.41).
369 Shared pathways were mostly assigned to carbohydrate metabolism; amino acid,
370 nucleotide, and fatty acid biosynthesis; and co-factor and vitamin biosynthesis (Fig. 6). The
371 microbiome bacteria contained unique pathways for organic carbon transportation and
372 degradation and vitamin B_{12} synthesis not found in *Microcystis*. Numerous pathways for organic
373 carbon degradation (D-galacturonate, D-glucuronate, galactose, glycogen, fatty acid, purine,
374 pyrimidine) and transportation (e.g., proline, maltose, galactose, maltose, mannose, D-xylose,
375 fructose, rhamnose, glycerol), as well as complete pathways for degradation of aromatics (mostly
376 anthropogenic pollutants, such as toluene, xylene, benzene, phthalate) were also identified in the
377 microbiome. Numerous important anaerobic bacterial pathways, including nitrogen fixation,
378 denitrification, and dissimilatory sulfate reduction were also detected in the *Microcystis*
379 microbiome, suggesting the anaerobic bacteria play an important role in nutrient cycling during
380 *Microcystis* blooms in these lakes. With the exception of some genes related to carbon cycling
381 (purine, pyrimidine, salicylate, and catechol degradation), as well as bacterial pathways involved
382 in methane metabolism (methane oxidation and formaldehyde assimilation), the majority of
383 potential biochemical function appeared to be associated with the 43 microbiome MAGs (Fig. 6,
384 Table S2).
385

Discussion

Genome reduction leads to a loss of function (Giovannoni et al. 2014), necessitating interactions with community members capable of carrying out those functions (Morris et al. 2012; Garcia et al. 2015). With small genomes compared to other algae (e.g., 4.2 Mbp vs 34.5 Mbp, Armbrust 2004; Gregory et al. 2007), *Microcystis* spp. are potentially missing some key metabolic functions (Steffen et al. 2012) and might be reliant on community members to fill in the metabolic gaps. Similar potential mutualist interactions have recently been mapped between *Microcystis* and their microbiome of associated bacteria (Xie et al. 2016; Li et al. 2018), corroborating the original idea of a 'phycosphere' of functional interaction within the *Microcystis* aggregates (Bell and Mitchell 1972; Paerl and Kellar 1978, 1979; Paerl and Millie 1996). Moreover, bacteria have been shown to complement algae in marine ecosystems by excreting large amounts of exometabolites including growth factors and biosynthetic precursors, as well as processing toxic metabolites (Morris et al. 2011; Pérez et al. 2016; Lee et al. 2017; Wienhausen et al. 2017). Given the predominance of such data across aquatic system, we hypothesized *M. aeruginosa* blooms have a microbiome of certain species of associated bacteria or perhaps metabolic functions that will be preserved across geographically distinct *Microcystis* blooms.

Microcystis aeruginosa was the most abundant *Microcystis* species across all lakes and this confirms that, at least at the level of 16S rRNA sequences, *M. aeruginosa* is a cosmopolitan bloom-forming species. We found remarkable phylogenetic relatedness among associated bacteria, and similar function between sites, despite those bacteria being taxonomically distinct at the 16S rRNA level. We found no relationship between community composition dissimilarity and geographic distance (Fig. 3a and 3b), indicating no distance-decay relationship as would be

409 expected for dispersal-limited species (Nekola and White 1999; Green and Bohannan 2006;
410 Nemergut et al. 2013). We also conclude that the functional potential of microbial communities
411 is more highly conserved than their taxonomic composition (Fig. 3A & Fig. 5). Similarly, Steffen
412 et al. (2012) found that the cyanobacterial bloom-associated bacterial communities across three
413 lakes (Erie, Taihu, St. Marys) differed taxonomically, while being functionally similar. Thus,
414 while OTU identity was variable across *Microcystis* microbiome bacteria, functional potential
415 appears to have been quite similar providing evidence that under bloom conditions, *M.*
416 *aeruginosa* is accompanied by a common suite of bacterial functionality, potentially forming an
417 interactome.

418 We found that *Microcystis* microbiome bacteria have functional potential not found in
419 *Microcystis,* and functional similarity was preserved globally. Pathways involving
420 photosynthesis (excluding the anoxygenic photosystem II pathway) and carbon fixation were
421 only detected in *Microcystis* (Fig. 6). The microbiome bacteria could potentially be contributing
422 to carbon recycling within the aggregates as many of the pathways found only in the bacteria are
423 related to carbohydrate breakdown (e.g., d-galacturonate, galactose and glycogen degradation)
424 and transport (e.g., numerous transport systems). The microbiome bacteria are likely tightly
425 associated with the carbon sources found in the *Microcystis* aggregates. Bacteria associated with
426 phytoplankton blooms have been found to utilize the carbon source glycolate from
427 phytoplankton (Lau et al. 2007; Paver and Kent 2010) and contain the glycolate oxidation
428 pathway. Unfortunately, the KEGG database does not contain this gene or related pathways thus
429 we were unable to detect it in our samples. Every metagenomic function database has its
430 shortcomings and for future studies we would recommend using multiple databases to cover as
431 many genes and pathways as possible. For example, using the COG database (clusters of

orthologous groups) to search for specific genes, we in fact found that the glcD gene, which is involved in the glycolate oxidation pathway, as well as the mycrocystin production genes (mcyA-I) were present in the microbiome and the *Microcystis*, respectively, in all of our lakes. In addition, we found that the microbiome potentially contributes methane metabolism pathways (methane oxidation and formaldehyde assimilation) which are used to convert methane into a useable carbon form. Recently, cyanobacteria have been suggested to produce methane during blooms (M. Bižić et al. unpublished preprint, doi: https://doi.org/10.1101/398958), although methane is typically produced by methanogenic archaea. While we did not analyze the archaea in our samples, other studies have found methanogenic archaea can be closely associated with cyanobacteria blooms (Batista et al. 2018), thus we would expect some methane production within aggregates.

In addition, the ethylmalonyl pathway was identified in the microbiome bacterial genomes. The ethylmalonyl pathway is a new acetate assimilation strategy in *Rhodobacter sphaeroides*, an anoxygenic phototrophic organism that lacks the key enzyme of the glyoxylate cycle, isocitrate lyase (Erb et al. 2007). Indeed, genes associated with anoxygenic photosystem II were identified in six microbiome MAGs (data not shown). Sulfur is a byproduct of anoxygenic photosynthesis and we also found evidence for thiosulfate oxidation and dissimilatory sulfate reduction pathways present in the microbiome bacteria. These results corroborate Li et al. (2018), who suggested that bacteria associated with *Microcystis* (termed epibionts by Li et al.) were essential for maintaining the redox balance and cycling different forms of sulfur within *Microcystis* aggregates. In addition, the vitamin B_{12} biosynthesis pathway, a necessary vitamin that *Microcystis* cannot produce, was only detected in the microbiome bacteria. Croft et al. (2005, 2006) proposed that most phytoplankton were likely auxotrophic for vitamin B_{12} and

455 other essential vitamins. Indeed, previous studies have also suggested that vitamin B_{12} was
456 provided to *Microcystis* by the associated bacteria and, moreover, that this relationship was
457 mutually beneficial for both groups (Xie et al. 2016; Li et al. 2018).

458 Most *Microcystis*-blooming lakes are typically recipients of urban and agricultural runoff.
459 Correspondingly, we found the *Microcystis* microbiome contained the degradation pathways for
460 many potentially harmful aromatic pollutants (e.g., benzene, benzoate, phthalate, etc.).
461 *Microcystis* could be benefiting from pollutant degradation as many of these aromatics have been
462 shown to inhibit phytoplankton growth (Häder and Gao 2015). Xi et al. (2016) also found a
463 group of aggregate-associated bacteria that contributed the whole benzoate degradation pathway
464 to the community, again pointing toward mutualism between the *Microcystis* and the *Microcystis*
465 microbiome. Together, these results support our hypothesis of a co-evolved interactome. We also
466 hypothesize that many of these microbiome functions (Fig. 6) are needed for growth and
467 dominance by *M. aeruginosa* during bloom conditions.

468 Loss of necessary metabolic functions is more common in bacterial communities than
469 previously thought (Hottes et al. 2013). Morris et al. (2012) found that the marine
470 cyanobacterium, *Prochlorococcus,* has lost many oxidative-stress genes and instead relies on
471 *Synechococcus*, another marine cyanobacterium, for removal of hydrogen peroxide from the
472 microenvironment. As such, *Prochlorococcus* may benefit from hydrogen peroxide removal by
473 way of *Synechococcus*, i.e., the "leaky" public good (*sensu* Morris et al. 2012), and the smaller
474 genome of *Prochlorococcus* affords it a selective advantage. A recent genomic study has shown
475 that *M. aeruginosa* also has a reduced genome with remarkable redundancy consisting of a set of
476 ~2,400 core genes and a large, variable pangenome, an additional set of genes unique to different
477 *M. aeruginosa* strains likely acquired through horizontal gene transfer (Humbert et al. 2013). We

speculate that the variability within the *M. aeruginosa* pangenome could be due to differences in the need for specific functions across different environments. Horizontal gene transfer and the Black Queen hypothesis provide two mechanisms for coping with changing environmental needs. "Leaky" functions may be lost from *Microcystis* when they are costly and a public good is available, as with the case of oxidative-stress genes in *Prochlorococcus*. However, when that public good becomes less predictable, those functions may be recouped through horizontal gene transfer if the new conditions provide a selective advantage to carriers for regaining the function. Steffen et al. (2012) corroborated this hypothesis as they found microbiome functional potential remained static between two *Microcystis* blooms. However, in one of the lakes (Taihu), *Microcystis* was reliant on Proteobacteria for nitrogen assimilation and metabolism, while in the other lakes (Erie and St. Marys), *Microcystis* carried out those functions. In the present study, we also found that the microbiome bacteria were potentially the sole contributors of nitrogen fixation, denitrification, and the urea cycle pathways (Fig. 6). *Microcystis* cannot fix nitrogen so this may indicate that *Microcystis* is relying on public goods across multiple blooms. This interactome of *M. aeruginosa* and its microbiome could be an ideal system to test for "Black Queen" functions.

Alternatively, the similarities we see in the microbiome bacterial community could be due in part to the *Microcystis* bloom creating specific habitats that are selective of certain types of heterotrophic bacteria. As previously mentioned, *Microcystis* blooms can change local environments, creating large amounts of particulate organic matter that have been shown to be an important nutrient source for shaping the bacterial community (e.g., Fogg G. E. and Pearsall William Harold 1952; Bell and Mitchell 1972; Paerl and Gallucci 1985). We saw through βNTI analysis that *Microcystis* microbiome communities were significantly phylogenetically related

(Fig. 4), with the negative value of βNTI indicating a common environmental filter across the lakes - *Microcystis* (sensu Dini-Andreote et al. 2015). Similarly, multiple studies (Yang et al. 2017; Xu et al. 2018; Shi et al. 2018) have shown that the *Microcystis*-associated (i.e., particle-attached, which is equivalent to our use on *Microcystis* microbiome) bacterial community composition appeared to be heavily structured by the bloom compared to free-living bacteria. Over the course of a *Microcystis* bloom, Parveen et al. (2013) found that *Microcystis* bloom aggregates provided habitat for a bacterial community distinct from the free-living bacteria. In addition, specific bacteria have also been found utilize parts of bloom aggregates. For example, bacteria in the genus, *Sphingomonas,* actively break down toxins while associated with *Microcystis* blooms (Dziallas and Grossart 2011). This bloom-as-habitat hypothesis could also explain the differences in attached and free-living bacteria described in previous studies (Yang et al. 2017; Xu et al. 2018; Shi et al. 2018). This hypothesis does not negate that *Microcystis* could be exchanging or receiving functions from these aggregate-associated bacteria. Additional metagenomic and multi-year bloom studies are needed to further parse these relationships.

The phylogenetic and functional similarities compared with the taxonomic dissimilarities we observed provide support to a growing body of evidence suggesting that community composition comparisons should be based on functional genes rather than strictly taxonomy (OTUs) (Burke et al. 2011; Oh et al. 2011). While clear similarities are shown using 16S rRNA genes, this method is taxonomically conservative and multiple different bacterial species could be represented by a single OTU. 16S rRNA genes represent a very small part of the genome, and evidence is growing which suggests that 16S rRNA is insufficient for distinguishing freshwater bacteria. For example, *Polynucleobacter* species, with 16S rRNA similarities ≥99%, have been shown to be distinct species based on whole genome comparisons and ecological isolation (Hahn

et al. 2016). In *Microcystis*, a threshold of 98-99% similarity is suggested to be sufficient in distinguishing some species (Harke et al. 2016). In our study, all *Microcystis* OTUs with greater than 1000 sequence abundance were assigned to *M. aeruginosa* at the species level, thus we are confident in this taxonomy, but also recognize the need for more detailed taxonomic analysis to further investigate the diversity of *Microcystis* within a bloom beyond the 97% similarity threshold for 16S rRNA genes.

In the current study, we have used 16S rRNA taxonomy paired with metagenomic functional analysis to describe the community composition and potential function of the associated bacterial community in nine geographically distinct *Microcystis* blooms. The multiple samples across continents has allowed us to confirm that *M. aeruginosa* is a cosmopolitan bloom former. The phylogenetic and functional similarity of associated bacteria across sites and the sets of complementary pathways found between the *Microcystis* and associated bacteria support the presence of a synergistic interactome. These results highlight the need for deeper investigation into *Microcystis* taxonomic identity and both *Microcystis* and microbiome functional capabilities at different times before, during, and after a bloom to fully elucidate the interactome relationship between *M. aeruginos*a and its microbiome. We conclude that the co-evolution within an interactome, i.e., between *M. aeruginosa* and its microbiome, could help explain the global distribution of, and dominance by, *M. aeruginosa* in diverse freshwater ecosystems.

Acknowledgments

We thank the members of the University of Oklahoma Plankton Ecology and Limnology Laboratory for helpful discussions, lab and statistical assistance, and for commenting on earlier

versions of this manuscript. In particular, we thank Jessica Beyer for invaluable statistical advice and critical editorial feedback and Stephen C. Cook for making Figure 1. Funding was provided by the University of Oklahoma Office of the Vice President for Research to L.R.K. and K.D.H., by the National Science Foundation (DEB 1831061) to K.D.H., L.R.K., H.W.P., M.M.S., and A.E.W., NSF OCE 1840715 and NIH-1P01ES028939-01to H.W.P, NSF DEB 1831106 to M.M.S., and through an NSF Graduate Research Fellowship to K.V.H. This work will be submitted by K.V.H. as partial fulfillment of the requirements for a Ph.D. degree from the University of Oklahoma.

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2

Armbrust, E. V. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science **306**: 79–86. doi:10.1126/science.1101156

Bankevich, A., and others. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19**: 455–477.

Batista, A., J. Woodhouse, H.-P. Grossart, and A. Giani. 2018. Methanogenic archaea associated to *Microcystis* sp. in field samples and in culture. Hydrobiologia 1–10.

Bell, W., and R. Mitchell. 1972. Chemotactic and growth responses of marine bacteria to algal extracellular products. Bio. Bull. **143**: 265–277. doi:10.2307/1540052

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120. doi:10.1093/bioinformatics/btu170

569 Bourne, D. G., P. G. Dennis, S. Uthicke, R. M. Soo, G. W. Tyson, and N. Webster. 2013. Coral
570 reef invertebrate microbiomes correlate with the presence of photosymbionts. ISME J. **7**:
571 1452–1458. doi:https://doi.org/10.1038/ismej.2012.172
572 Burke, C., P. Steinberg, D. Rusch, S. Kjelleberg, and T. Thomas. 2011. Bacterial community
573 assembly based on functional genes rather than species. PNAS **108**: 14288–14293.
574 doi:10.1073/pnas.1101591108
575 Cai, H., H. Jiang, L. R. Krumholz, and Z. Yang. 2014. Bacterial community composition of size-
576 fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic
577 freshwater lake. PLoS One **9**: e102879. doi:10.1371/journal.pone.0102879
578 Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen, and R. Knight.
579 2009. PyNAST: a flexible tool for aligning sequences to a template alignment.
580 Bioinformatics **26**: 266–267. doi:10.1093/bioinformatics/btp636
581 Caporaso, J. G., and others. 2010. QIIME allows analysis of high-throughput community
582 sequencing data. Nat. Methods **7**: 335–336. doi:10.1038/nmeth.f.303
583 Carey, C. C., B. W. Ibelings, E. P. Hoffmann, D. P. Hamilton, and J. D. Brookes. 2012. Eco-
584 physiological adaptations that favour freshwater cyanobacteria in a changing climate.
585 Water Res. **46**: 1394–1407. doi:10.1016/j.watres.2011.12.016
586 Carmichael, W. W. 2001. Health Effects of Toxin-Producing Cyanobacteria: "The
587 CyanoHABs." Hum. Ecol. Risk. Assess. **7**: 1393–1407. doi:10.1080/20018091095087
588 Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C.
589 Lancelot, and G. E. Likens. 2009. Controlling eutrophication: Nitrogen and phosphorus.
590 Science **323**: 1014–1015. doi:10.1126/science.1167755

591 Croft, M. T., A. D. Lawrence, E. Raux-Deery, M. J. Warren, and A. G. Smith. 2005. Algae
592 acquire vitamin B12 through a symbiotic relationship with bacteria. Nature **438**: 90–93.
593 doi:10.1038/nature04056
594 Croft, M. T., M. J. Warren, and A. G. Smith. 2006. Algae need their vitamins. Eukaryot. Cell **5**:
595 1175–1183. doi:10.1128/EC.00097-06
596 Dini-Andreote, F., J. C. Stegen, J. D. van Elsas, and J. F. Salles. 2015. Disentangling
597 mechanisms that mediate the balance between stochastic and deterministic processes in
598 microbial succession. PNAS **112**: E1326–E1332. doi:10.1073/pnas.1414261112
599 Dziallas, C., and H. Grossart. 2011. Temperature and biotic factors influence bacterial
600 communities associated with the cyanobacterium *Microcystis* sp. Environ. Microbiol. **13**:
601 1632–1641. dio:10.1111/j.1462-2920.2011.02479.x
602 Dziallas, C., and H.-P. Grossart. 2012. Microbial interactions with the cyanobacterium
603 *Microcystis aeruginosa* and their dependence on temperature. Mar. Biol. **159**: 2389–
604 2398. doi:10.1007/s00227-012-1927-4
605 Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics
606 **26**: 2460–2461. doi:10.1093/bioinformatics/btq461
607 Edgar, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads.
608 Nature Methods **10**: 996–998. doi:10.1038/nmeth.2604
609 Eiler, A., and S. Bertilsson. 2004. Composition of freshwater bacterial communities associated
610 with cyanobacterial blooms in four Swedish lakes. Environ. Mircobiol. **6**: 1228–1243.
611 doi:10.1111/j.1462-2920.2004.00657.x

612 Erb, T. J., I. A. Berg, V. Brecht, M. Muller, G. Fuchs, and B. E. Alber. 2007. Synthesis of C5-
613 dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The
614 ethylmalonyl-CoA pathway. PNAS **104**: 10631–10636. doi:10.1073/pnas.0702791104
615 Fierer, N. 2017. Embracing the unknown: Disentangling the complexities of the soil microbiome.
616 Nat. Rev. Microbiol. **15**: 579–590. doi:10.1038/nrmicro.2017.87
617 Fogg G. E., and Pearsall William Harold. 1952. The production of extracellular nitrogenous
618 substances by a blue-green alga. Proc. R. Soc. B **139**: 372–397.
619 doi:10.1098/rspb.1952.0019
620 Garcia, S. L., M. Buck, K. D. McMahon, H.-P. Grossart, A. Eiler, and F. Warnecke. 2015.
621 Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated
622 freshwater model community. Mol. Ecol. **24**: 4449–4459. doi:10.1111/mec.13319
623 Giovannoni, S. J., J. Cameron Thrash, and B. Temperton. 2014. Implications of streamlining
624 theory for microbial ecology. ISME J. **8**: 1553–1565. doi:10.1038/ismej.2014.60
625 Green, J., and B. J. M. Bohannan. 2006. Spatial scaling of microbial biodiversity. Trends Ecol.
626 Evol. **21**: 501–507. doi:10.1016/j.tree.2006.06.012
627 Gregory, T. R., and others. 2007. Eukaryotic genome size databases. Nucleic Acids Res. **35**:
628 D332–D338. doi:10.1093/nar/gkl828
629 Hahn, M. W., J. Jezberová, U. Koll, T. Saueressig-Beck, and J. Schmidt. 2016. Complete
630 ecological isolation and cryptic diversity in *Polynucleobacter* bacteria not resolved by
631 16S rRNA gene sequences. ISME J. **10**: 1642–1655. doi:10.1038/ismej.2015.237
632 Harke, M. J., M. M. Steffen, C. J. Gobler, T. G. Otten, S. W. Wilhelm, S. A. Wood, and H. W.
633 Paerl. 2016. A review of the global ecology, genomics, and biogeography of the toxic

cyanobacterium, *Microcystis* spp. Harmful Algae **54**: 4–20. doi:10.1016/j.hal.2015.12.007

Havens, K. E. 2008. Cyanobacteria blooms: Effects on aquatic ecosystems, p. 733–747. *In* H.K. Hudnell [ed.], Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. Springer New York.

Herdman, M., M. Janvier, R. Rippka, and R. Y. Stanier. 1979. Genome size of cyanobacteria. J. Gen. Microbiol. **111**: 73–85. doi:10.1099/00221287-111-1-73

Hottes, A. K., P. L. Freddolino, A. Khare, Z. N. Donnell, J. C. Liu, and S. Tavazoie. 2013. Bacterial adaptation through loss of function. PLOS Genet. **9**: e1003617. doi:10.1371/journal.pgen.1003617

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471. doi:10.1038/s41579-018-0040-1

Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature **486**: 207–214. doi:10.1038/nature11234

Humbert, J.-F., and others. 2013. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. PLoS One **8**: e70747. doi:10.1371/journal.pone.0070747

Kanehisa, M., Y. Sato, and K. Morishima. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. **428**: 726–731. doi:10.1016/j.jmb.2015.11.006

Kang, D. D., J. Froula, R. Egan, and Z. Wang. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ **3**: e1165.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**: 1463–1464.

Klindworth, A., E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F. O. Glöckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. **41**: e1–e1. doi:10.1093/nar/gks808

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**: 357. doi:10.1038/nmeth.1923

Lau, W. W., R. G. Keil, and E. V. Armbrust. 2007. Succession and diel transcriptional response of the glycolate-utilizing component of the bacterial community during a spring phytoplankton bloom. Appl. Environ. Microbiol. **73**: 2440–2450.

Lee, M. D., N. G. Walworth, E. L. McParland, F.-X. Fu, T. J. Mincer, N. M. Levine, D. A. Hutchins, and E. A. Webb. 2017. The Trichodesmium consortium: conserved heterotrophic co-occurrence and genomic signatures of potential interactions. The ISME Journal **11**: 1813.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**: 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., P. Xing, M. Chen, Y. Bian, and Q. L. Wu. 2011. Short-term bacterial community composition dynamics in response to accumulation and breakdown of *Microcystis* blooms. Water Res. **45**: 1702–1710. doi:10.1016/j.watres.2010.11.011

Li, Q., and others. 2018. A large-scale comparative metagenomic study reveals the functional interactions in six bloom-forming *Microcystis*-epibiont communities. Front. Microbiol. **9**. doi:10.3389/fmicb.2018.00746

MacArthur, R. H. 1984. Geographical ecology: patterns in the distribution of species, Princeton University Press.

Mantzouki, E., and others. 2018. Temperature effects explain continental scale distribution of cyanobacterial toxins. Toxins **10**: 156. doi:10.3390/toxins10040156

Morris, J. J., Z. I. Johnson, M. J. Szul, M. Keller, and E. R. Zinser. 2011. Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. PLoS One **6**: e16805. doi:10.1371/journal.pone.0016805

Morris, J. J., R. E. Lenski, and E. R. Zinser. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio **3**: e00036-12.

Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. J. Biogeogr. **26**: 867–878. doi:10.1046/j.1365-2699.1999.00305.x

Nemergut, D. R., and others. 2013. Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. **77**: 342–356. doi:10.1128/MMBR.00051-12

Oh, S., A. Caro-Quintero, D. Tsementzi, N. DeLeon-Rodriguez, C. Luo, R. Poretsky, and K. T. Konstantinidis. 2011. Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. Appl. Environ. Microbiol. **77**: 6000–6011. doi:10.1128/AEM.00107-11

Oksanen, J., and others. 2013. vegan: Community ecology package. R Package version 2.5-3. Available from http://CRAN.R-project.org/package=vegan.

O'Neil, J. M., T. W. Davis, M. A. Burford, and C. J. Gobler. 2012. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. Harmful Algae **14**: 313–334. doi:10.1016/j.hal.2011.10.027

Paerl, H. W., and K. K. Gallucci. 1985. Role of chemotaxis in establishing a specific nitrogen-fixing cyanobacterial-bacterial association. Science **227**: 647–649. doi:10.1126/science.227.4687.647

Paerl, H. W., and P. E. Kellar. 1978. Significance of bacterial-*Anabaena* (cyanophyceae) associations with respect to N_2 fixation in freshwater. J. Phycol. **14**: 254–260. doi:10.1111/j.1529-8817.1978.tb00295.x

Paerl, H. W., and P. E. Kellar. 1979. Nitrogen-fixing *Anabaena*: Physiological adaptations instrumental in maintaining surface blooms. Science **204**: 620–622. doi:10.1126/science.204.4393.620

Paerl, H. W., and D. F. Millie. 1996. Physiological ecology of toxic aquatic cyanobacteria. Phycologia **35**: 160–167. doi:10.2216/i0031-8884-35-6S-160.1

Paerl, H. W., and T. G. Otten. 2013. Harmful cyanobacterial blooms: Causes, consequences, and controls. Microb. Ecol. **65**: 995–1010. doi:10.1007/s00248-012-0159-y

Paerl, H. W., and T. G. Otten. 2016. Duelling 'CyanoHABs': Unravelling the environmental drivers controlling dominance and succession among diazotrophic and non- N2- fixing harmful cyanobacteria. Environ. Mircobiol. **18**: 316–324. doi:10.1111/1462-2920.13035

Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. **46**: 1349–1363. doi:10.1016/j.watres.2011.08.002

721 Paerl, H. W., and others. 2016. It takes two to tango: When and where dual nutrient (N & P)
722 reductions are needed to protect lakes and downstream ecosystems. Environ. Sci.
723 Technol. **50**: 10805–10813.
724 Paerl, H. W., H. Xu, M. J. McCarthy, G. Zhu, B. Qin, Y. Li, and W. S. Gardner. 2011.
725 Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu,
726 China): The need for a dual nutrient (N & P) management strategy. Water Res. **45**: 1973–
727 1983. doi:10.1016/j.watres.2010.09.018
728 Pande, S., and C. Kost. 2017. Bacterial unculturability and the formation of intercellular
729 metabolic networks. Trends Microbiol. **25**: 349–361. doi:10.1016/j.tim.2017.02.015
730 Parveen, B., V. Ravet, C. Djediat, I. Mary, C. Quiblier, D. Debroas, and J.-F. Humbert. 2013.
731 Bacterial communities associated with *Microcystis* colonies differ from free-living
732 communities living in the same ecosystem: Bacterial diversity inside *Microcystis*
733 colonies. Environ. Mircobiol. Rep. n/a-n/a. doi:10.1111/1758-2229.12071
734 Paver, S. F., and A. D. Kent. 2010. Temporal patterns in glycolate-utilizing bacterial community
735 composition correlate with phytoplankton population dynamics in humic lakes. Microb.
736 Ecol. **60**: 406–418. doi:10.1007/s00248-010-9722-6
737 Pérez, A. A., D. A. Rodionov, and D. A. Bryant. 2016. Identification and regulation of genes for
738 cobalamin transport in the cyanobacterium *Synechococcus* sp. strain PCC 7002. J.
739 Bacteriol. **198**: 2753–2761. doi:10.1128/JB.00476-16
740 Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2–approximately maximum-likelihood
741 trees for large alignments. PLoS One **5**: e9490.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2012. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. **41**: D590–D596. doi:10.1093/nar/gks1219

R Development Core Team. 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Salomon, P. S., S. Janson, and E. Granéli. 2003. Molecular identification of bacteria associated with filaments of *Nodularia spumigena* and their effect on the cyanobacterial growth. Harmful Algae **2**: 261–272. doi:10.1016/S1568-9883(03)00045-3

Schindler, D. W. 2012. The dilemma of controlling cultural eutrophication of lakes. Proc. R. Soc. B **279**: 4322–4333. doi:10.1098/rspb.2012.1032

Schindler, D. W., S. R. Carpenter, S. C. Chapra, R. E. Hecky, and D. M. Orihel. 2016. Reducing phosphorus to curb lake eutrophication is a success. Environ. Sci. Technol. **50**: 8923–8929. doi:10.1021/acs.est.6b02204

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**: 2068–2069. doi:10.1093/bioinformatics/btu153

Shi, L., and others. 2018. Large buoyant particles dominated by cyanobacterial colonies harbor distinct bacterial communities from small suspended particles and free- living bacteria in the water column. MicrobiologyOpen **7**: e00608. doi:10.1002/mbo3.608

Sieber, C. M., A. J. Probst, A. Sharrar, B. C. Thomas, M. Hess, S. G. Tringe, and J. F. Banfield. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. **3**: 836. doi:10.1038/s41564-018-0171-1

Steffen, M. M., Z. Li, T. C. Effler, L. J. Hauser, G. L. Boyer, and S. W. Wilhelm. 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on

765 two continents S. Bertilsson [ed.]. PLoS One **7**: e44002.
766 doi:10.1371/journal.pone.0044002
767 Stegen, J. C., X. Lin, A. E. Konopka, and J. K. Fredrickson. 2012. Stochastic and deterministic
768 assembly processes in subsurface microbial communities. ISME J. **6**: 1653–1664.
769 doi:10.1038/ismej.2012.22
770 Steiner, K., S. A. Wood, J. Puddick, I. Hawes, D. R. Dietrich, and D. P. Hamilton. 2017. A
771 comparison of bacterial community structure, activity and microcystins associated with
772 formation and breakdown of a cyanobacterial scum. Aquat. Microb. Ecol. **80**: 243–256.
773 doi:10.3354/ame01852
774 Swan, B. K., and others. 2013. Prevalent genome streamlining and latitudinal divergence of
775 planktonic bacteria in the surface ocean. PNAS **110**: 11463–11468.
776 doi:10.1073/pnas.1304246110
777 Swenson, N. G. 2014. Functional and phylogenetic ecology in R, Springer.
778 Uritskiy, G. V., J. DiRuggiero, and J. Taylor. 2018. MetaWRAP—a flexible pipeline for
779 genome-resolved metagenomic data analysis. Microbiome **6**: 158. doi:10.1186/s40168-
780 018-0541-1
781 Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid
782 assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ.
783 Microbiol. **73**: 5261–5267. doi:10.1128/AEM.00062-07
784 Wawrik, B., and others. 2012. Field and laboratory studies on the bioconversion of coal to
785 methane in the San Juan Basin. FEMS Microbiol. Ecol. **81**: 26–42. doi:10.1111/j.1574-
786 6941.2011.01272.x

787 Webb, C. O. 2000. Exploring the Phylogenetic Structure of Ecological Communities: An
788 Example for Rain Forest Trees. Am. Nat. **156**: 145–155. doi:10.1086/303378
789 Wienhausen, G., B. E. Noriega-Ortega, J. Niggemann, T. Dittmar, and M. Simon. 2017. The
790 exometabolome of two model strains of the *Roseobacter* group: A marketplace of
791 microbial metabolites. Front. Microbiol. **8**. doi:10.3389/fmicb.2017.01985
792 Wu, Y.-W., B. A. Simmons, and S. W. Singer. 2015. MaxBin 2.0: An automated binning
793 algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics **32**:
794 605–607. doi:10.1093/bioinformatics/btv638
795 Xie, M., M. Ren, C. Yang, H. Yi, Z. Li, T. Li, and J. Zhao. 2016. Metagenomic analysis reveals
796 symbiotic relationship among bacteria in *Microcystis*-dominated community. Front.
797 Microbiol. **7**. doi:10.3389/fmicb.2016.00056
798 Xu, H., and others. 2018. Contrasting network features between free-living and particle-attached
799 bacterial communities in Taihu Lake. Microb. Ecol. **76**: 303–313. doi:10.1007/s00248-
800 017-1131-7
801 Yang, C., and others. 2017. Distinct network interactions in particle-associated and free-living
802 bacterial communities during a *Microcystis aeruginosa* bloom in a Plateau Lake. Front.
803 Microbiol. **8**. doi:10.3389/fmicb.2017.01202
804

805 ***FIGURE CAPTIONS:***

806 **Figure 1.** Location of the 12 lakes across the globe. These samples represent a 280° longitudinal
807 and 90° latitudinal gradient.

808 **Figure 2.** (A) Relative abundance of Bacteria classes in the nine lakes. The lakes are arranged in
809 order from left to right of increasing percent *Microcystis* in the community. Classes less than
810 1% of the total relative abundance were grouped together as a single group denoted "<1%
811 abund". Oxyphotobacteria (cyanobacteria) were split into two groups: *Microcystis* only in
812 one and all other cyanobacteria in the second. (B) Relative abundance of non-*Microcystis*
813 (i.e., microbiome) bacterial classes.

814 **Figure 3.** Scatter plots of community dissimilarity in the microbiome as related to geographic
815 distance. (A) The non-significant (GLM Deviance explained = 3.7%, p = 0.2) relationship
816 between taxonomic Bray-Curtis dissimilarity and geographic distance where the higher Bray-
817 Curtis values indicate fewer species in common between sites. (B) Abundance weighted
818 UniFrac did not scale significantly by geographic distance (GLM DE = 0.82%, p = 0.55).
819 Here higher values of UniFrac indicate there is little overlap in species between communities
820 whereas lower values indicate the communities are more similar.

821 **Figure 4.** Distributions of *within* community phylogenetic relatedness (αNTI -Nearest Taxon
822 Index), and the phylogenetic relatedness *between* two communities (βNTI) of the nine
823 sampled lakes. Values below -2 or above +2 standard deviations from the null (indicated by
824 the red rectangle) are statistically significantly different from random. Black dashed lines
825 indicate the mean of the observed distributions. The mean of the αNTI distribution is 4.64
826 and the mean of the βNTI distribution is -3.58. αNTI is a measure of community
827 phylogenetic structure and relatedness, where positive deviations from the null expectation

indicate the species in the community are more phylogenetically related (clustered) than expected by chance (as seen here), and negative deviations indicate the species are more phylogenetically distant (overdispersed). The observed αNTI was significantly different from the null ($t = 10.13$, $p < 0.001$). βNTI measures phylogenetic relatedness between two communities with values greater than the null meaning lower relatedness than expected by chance and values lower than the null meaning higher relatedness than expected by chance (as seen here). Our βNTI is significantly different from random ($t = -12.65$, $p < 0.001$).

Figure 5. The dissimilarity between the *Microcystis* microbiome community's metagenomic function was not significantly correlated with geographic distance (GLM DE = 1.97%, $p = 0.41$) and was overall low (low values of Bray-Curtis dissimilarity).

Figure 6. Venn diagram showing the distribution of complete or nearly complete (no more than one gene missing) KEGG modules in *Microcystis* and the microbiome bacteria. See Supplementary Table 2 for details and indication for involvment in major elemental cycling. KEGG modules in bold print with astricks were not detected in *Microcystis* and the microbiome bacterial MAGs.

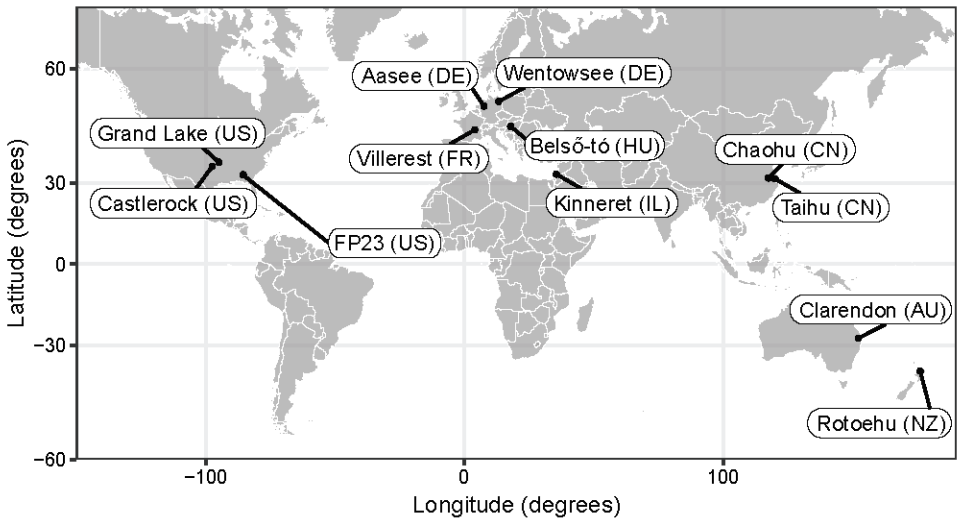


Figure 1. Location of the 12 lakes across the globe. These samples represent a 280° longitudinal and 90° latitudinal gradient.

Dimensions: 12.7 cm wide by 6.6 cm tall

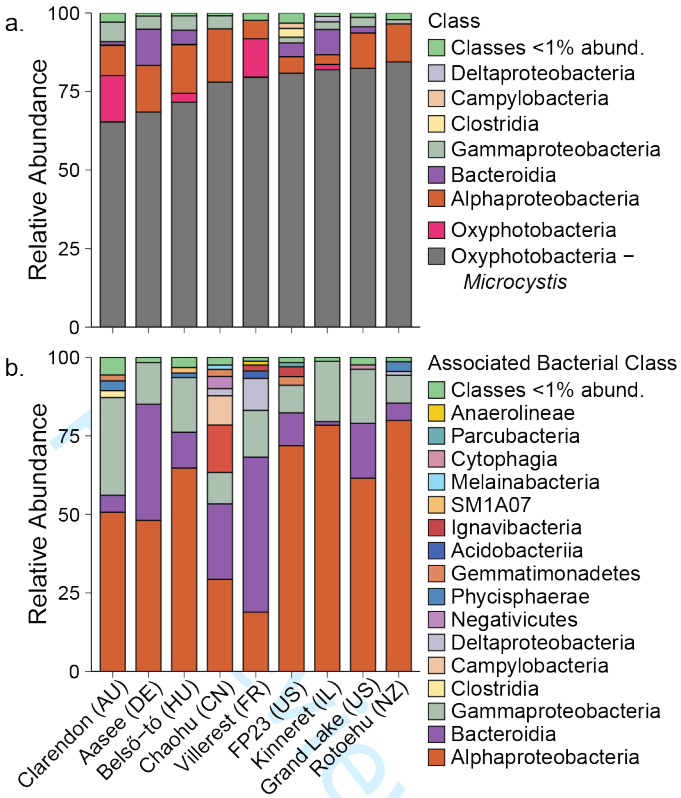


Figure 2. (A) Relative abundance of Bacteria classes in the nine lakes. The lakes are arranged in order from left to right of increasing percent *Microcystis* in the community. Classes less than 1% of the total relative abundance were grouped together as a single group denoted "<1% abund". Oxyphotobacteria (cyanobacteria) were split into two groups: *Microcystis* only in one and all other cyanobacteria in the second. (B) Relative abundance of non-*Microcystis* (i.e., microbiome) bacterial classes.

Dimensions: 8.9 cm wide by 10.7 cm high

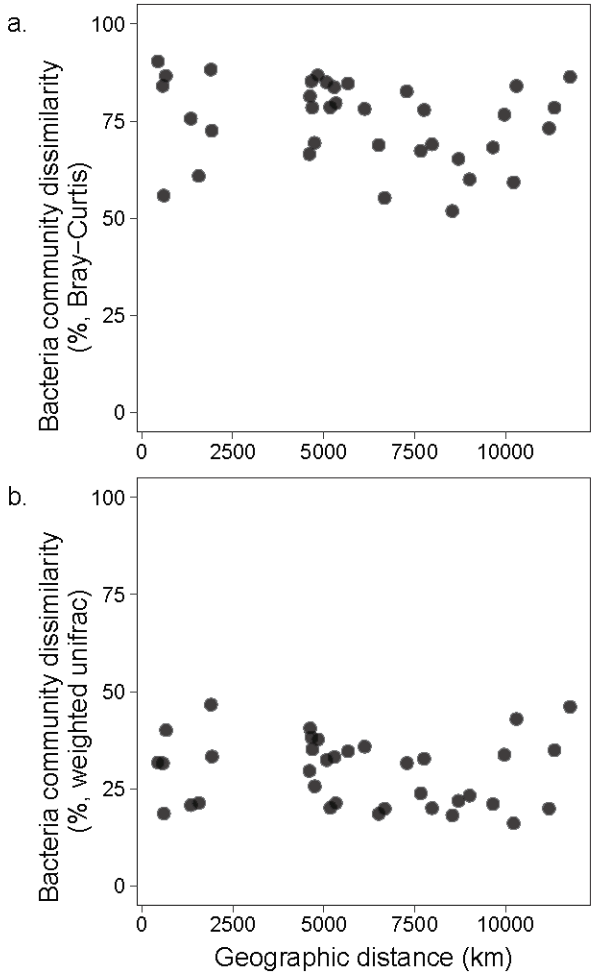


Figure 3. Scatter plots of community dissimilarity in the microbiome as related to geographic distance. (A) The non-significant (GLM Deviance explained = 3.7%, p = 0.2) relationship between taxonomic Bray-Curtis dissimilarity and geographic distance where the higher Bray-Curtis values indicate fewer species in common between sites. (B) Abundance weighted UniFrac did not scale significantly by geographic distance (GLM DE = 0.82%, p = 0.55). Here higher values of UniFrac indicate there is little overlap in species between communities whereas lower values indicate the communities are more similar.

Dimensions: 7.6 cm wide by 12.7 cm tall

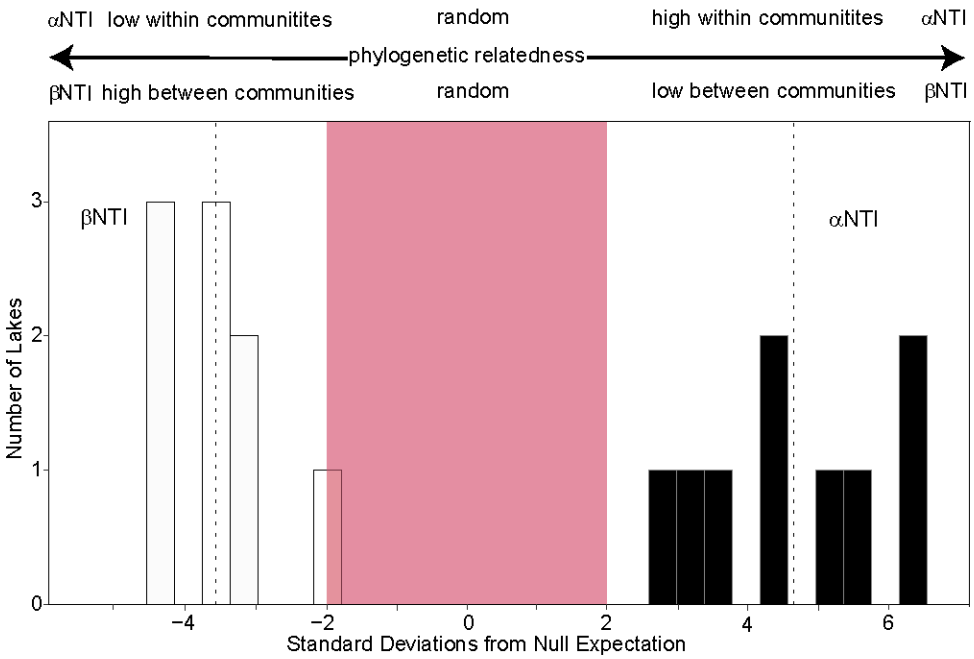


Figure 4. Distributions of within community phylogenetic relatedness (αNTI -Nearest Taxon Index), and the phylogenetic relatedness between two communities (βNTI) of the nine sampled lakes. Values below -2 or above +2 standard deviations from the null (indicated by the red rectangle) are statistically significantly different from random. Black dashed lines indicate the mean of the observed distributions. The mean of the αNTI distribution is 4.64 and the mean of the βNTI distribution is -3.58. αNTI is a measure of community phylogenetic structure and relatedness, where positive deviations from the null expectation indicate the species in the community are more phylogenetically related (clustered) than expected by chance (as seen here), and negative deviations indicate the species are more phylogenetically distant (overdispersed). The observed αNTI was significantly different from the null ($t = 10.13$, $p < 0.001$). βNTI measures phylogenetic relatedness between two communities with values greater than the null meaning lower relatedness than expected by chance and values lower than the null meaning higher relatedness than expected by chance (as seen here). Our βNTI is significantly different from random ($t = -12.65$, $p < 0.001$).

Dimensions: 12.7 cm wide by 8.4 cm tall

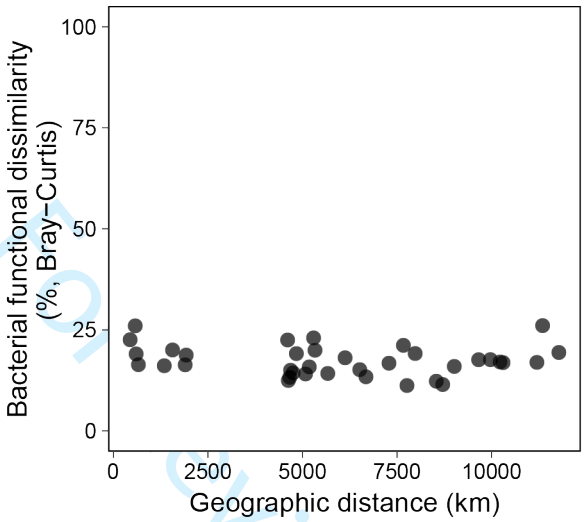


Figure 5. The dissimilarity between the *Microcystis* microbiome community's metagenomic function was not significantly correlated with geographic distance (GLM DE = 1.97%, p = 0.41) and was overall low (low values of Bray-Curtis dissimilarity).

Dimensions: 7.62 cm wide by 6.9 cm tall

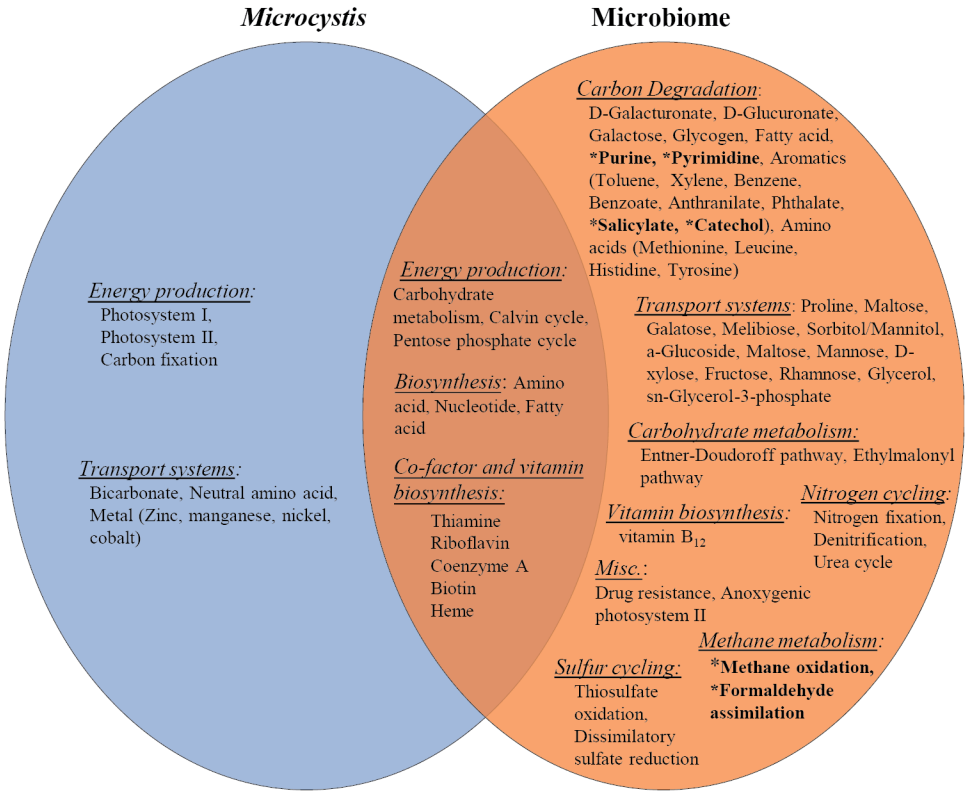


Figure 6. Venn diagram showing the distribution of complete or nearly complete (no more than one gene missing) KEGG modules in *Microcystis* and the microbiome bacteria. See Supplementary Table 2 for details and indication for involvment in major elemental cycling. KEGG modules in bold print with astricks were not detected in *Microcystis* and the microbiome bacterial MAGs.

Dimensions: 12.7 cm wide by 11.43 cm tall

Supporting Information:

The global *Microcystis* interactome

Hooker, K.V.,[1,2] C. Li,[3], H. Cai[1], L.R. Krumholz,[3] K.D. Hambright,[1,2,*] H.W. Paerl,[4] M.M. Steffen,[5] A.E. Wilson,[6] M. Burford,[7] H.-P. Grossart,[8] D. P. Hamilton,[7,9] H. Jiang,[10] A. Sukenik,[11] D. Latour,[12] E.I. Meyer,[13] J. Padisák,[14] B. Qin,[10] R.M. Zamor,[15,a] and G. Zhu[10]

Table S1. Summary of metagenome data for each lake.

Lake	category	contigs number	contigs bases(Mbp)	N50(bp)	Max length(bp)	protein-coding gene number
Belső-tó	Bacteria	33496	27.1	765	74324	26699
	Microcystis	890	4.7	16388	81696	4267
Chaohu	Bacteria	23682	19.8	794	36145	18385
	Microcystis	5258	7.8	1944	54606	6268
Clarendon	Bacteria	82636	104.8	1392	321729	102844
	Microcystis	6740	10.9	2336	38246	9427
FP23	Bacteria	155545	19.5	1454	49139	28976
	Microcystis	4816	7.9	2248	35883	6825
Grand	Bacteria	73551	92.3	1418	105344	89141
	Microcystis	9312	11.1	1418	57431	8881
Kinneret	Bacteria	26202	37.0	1817	101729	37307
	Microcystis	4943	9.1	2627	55640	8028
Aasee	Bacteria	36892	57.9	2441	299862	56457
	Microcystis	1122	5.5	19101	77987	5191
Rotoehu	Bacteria	26009	31.1	1368	41132	32313
	Microcystis	503	4.3	16710	49814	4098
Villerest	Bacteria	30454	40.5	1691	97662	36391
	Microcystis	2619	6.7	5702	54900	5927

Table S2. KEGG orthology numbers and functional pathways found in *Microcystis* and the microbiome bacteria, indicating involvement in C, N, P, or S cycling. Complete pathways are indicated by black fill ■, while modules with no more than one pathway missing, are indicated by blue fill ■. Empty cells indicate missing or partial (>1 missing) pathways.

Pathway	*Microcystis*	Bacteria	Biogeo-chemical cycle
Carbohydrate metabolism			
M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate			**C**
M00002 Glycolysis, core module involving three-carbon compounds			**C**
M00003 Gluconeogenesis, oxaloacetate => fructose-6P			**C, P**
M00307 Pyruvate oxidation, pyruvate => acetyl-CoA			**C**
M00009 Citrate cycle (TCA cycle, Krebs cycle)			**C**
M00010 Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate			**C, S**
M00011 Citrate cycle, second carbon oxidation, 2-oxoglutarate => oxaloacetate			**C**
M00004 Pentose phosphate pathway (Pentose phosphate cycle)			**C, P**
M00006 Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P			**C,P**
M00007 Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P			**C,P**
M00580 Pentose phosphate pathway, archaea, fructose 6P => ribose 5P			**C,P**
M00005 PRPP biosynthesis, ribose 5P => PRPP			**C,P**
M00008 Entner-Doudoroff pathway, glucose-6P => glyceraldehyde-3P + pyruvate			**C,P**
M00308 Semi-phosphorylative Entner-Doudoroff pathway, gluconate => glycerate-3P			**C,P**

M00309 Non-phosphorylative Entner-Doudoroff pathway, gluconate/galactonate => glycerate	**C**
M00854 Glycogen biosynthesis, glucose-1P => glycogen/starch	**C,P**
M00855 Glycogen degradation, glycogen => glucose-6P	**C,P**
M00565 Trehalose biosynthesis, D-glucose 1P => trehalose	**C,P**
M00549 Nucleotide sugar biosynthesis, glucose => UDP-glucose	**C,P**
M00631 D-Galacturonate degradation, D-galacturonate => pyruvate + D glyceraldehyde 3P	**C,P**
M00061 D-Glucuronate degradation, D-glucuronate => pyruvate + D-glyceraldehyde 3P	**C,P**
M00632 Galactose degradation, Leloir pathway, galactose => alpha-D-glucose-1P	**C,P**
M00552 D-galactonate degradation, De Ley-Doudoroff pathway, D-galactonate => glycerate-3P	**C,P**
M00554 Nucleotide sugar biosynthesis, galactose => UDP-galactose	**C,P**
M00012 Glyoxylate cycle	**C**
M00373 Ethylmalonyl pathway	**C**
M00013 Malonate semialdehyde pathway, propanoyl-CoA => acetyl-CoA	**C**
M00741 Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA	**C**
M00130 Inositol phosphate metabolism, PI=> PIP2 => Ins(1,4,5)P3 => Ins(1,3,4,5)P4	**C,P**
M00132 Inositol phosphate metabolism, Ins(1,3,4)P3 => phytate	**C**

Energy metabolism

Carbon fixation

M00165 Reductive pentose phosphate cycle (Calvin cycle)	**C**
M00166 Reductive pentose phosphate cycle, ribulose-5P => glyceraldehyde-3P	**C,P**
M00167 Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P	**C,P**

M00168 CAM (Crassulacean acid metabolism), dark	**C**
M00173 Reductive citrate cycle (Arnon-Buchanan cycle)	**C**
M00579 Phosphate acetyltransferase-acetate kinase pathway, acetyl-CoA => acetate	**C,P**
Methane metabolism	
M00345 Formaldehyde assimilation, ribulose monophosphate pathway	**C**
M00358 Coenzyme M biosynthesis	**C**
M00174 Methane oxidation, methanotroph, methane => formaldehyde	**C**
M00346 Formaldehyde assimilation, serine pathway	**C**
M00344 Formaldehyde assimilation, xylulose monophosphate pathway	**C,P**
M00378 F420 biosynthesis	**C**
Nitrogen metabolism	
M00531 Assimilatory nitrate reduction, nitrate => ammonia	**N**
M00175 Nitrogen fixation, nitrogen => ammonia	**N**
M00530 Dissimilatory nitrate reduction, nitrate => ammonia	**N**
M00529 Denitrification, nitrate => nitrogen	**N**
M00528 Nitrification, ammonia => nitrite	**N**
M00804 Complete nitrification, comammox, ammonia => nitrite => nitrate	**N**
Sulfur metabolism	
M00176 Assimilatory sulfate reduction, sulfate => H2S	S
M00595 Thiosulfate oxidation by SOX complex, thiosulfate => sulfate	**S**
M00596 Dissimilatory sulfate reduction, sulfate => H2S	**S**

Photosynthesis

M00161 Photosystem II C

M00163 Photosystem I C

M00597 Anoxygenic photosystem II **C**

ATP synthesis (Structural complex)

M00144 NADH:quinone oxidoreductase, prokaryotes

M00146 NADH dehydrogenase (ubiquinone) 1 alpha subcomplex

M00145 NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria

M00149 Succinate dehydrogenase, prokaryotes

M00162 Cytochrome b6f complex

M00154 Cytochrome c oxidase

M00155 Cytochrome c oxidase, prokaryotes

M00153 Cytochrome bd ubiquinol oxidase

M00157 F-type ATPase, prokaryotes and chloroplasts

M00150 Fumarate reductase, prokaryotes

M00162 Cytochrome b6f complex

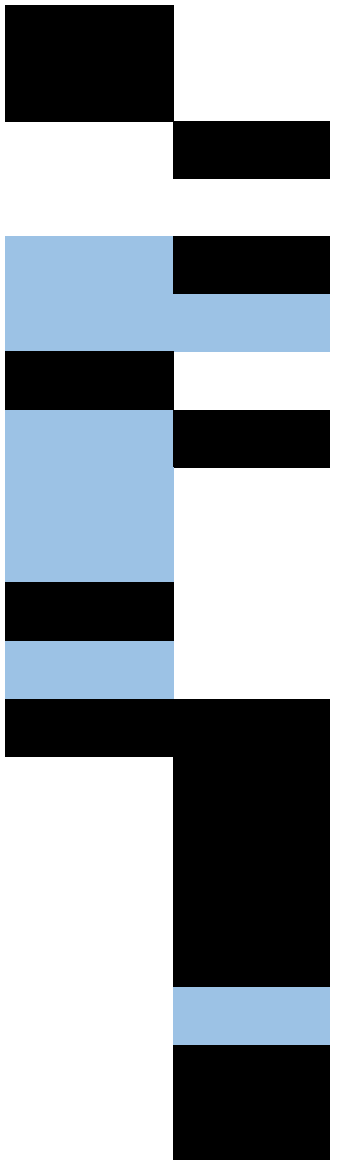
M00151 Cytochrome bc1 complex respiratory unit

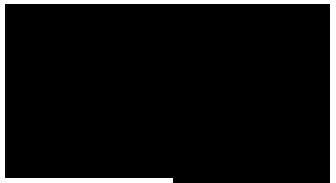
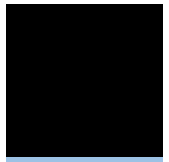
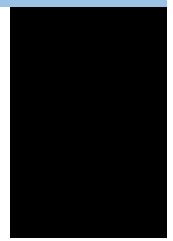
M00152 Cytochrome bc1 complex

M00154 Cytochrome c oxidase

M00155 Cytochrome c oxidase, prokaryotes

M00153 Cytochrome bd ubiquinol oxidase



M00417 Cytochrome o ubiquinol oxidase		
M00156 Cytochrome c oxidase, cbb3-type		
M00159 V-type ATPase, prokaryotes		
Lipid metabolism		
Fatty acid metabolism		
M00082 Fatty acid biosynthesis, initiation		**C**
M00083 Fatty acid biosynthesis, elongation		**C**
M00086 beta-Oxidation, acyl-CoA synthesis		**C**
M00087 beta-Oxidation		**C**
M00861 beta-Oxidation, peroxisome, VLCFA		**C**
Sterol biosynthesis		
M00862 beta-Oxidation, peroxisome, tri/dihydroxycholestanoyl-CoA => choloyl/chenodeoxycholoyl-CoA		**C**
M00107 Steroid hormone biosynthesis, cholesterol => prognenolone => progesterone		**C**
M00110 C19/C18-Steroid hormone biosynthesis, pregnenolone => androstenedione => estrone		**C**
Lipid metabolism		
M00098 Acylglycerol degradation		C
M00088 Ketone body biosynthesis, acetyl-CoA => acetoacetate/3-hydroxybutyrate/acetone		**C**
M00089 Triacylglycerol biosynthesis		**C**
M00090 Phosphatidylcholine (PC) biosynthesis, choline => PC		**C,P**
M00091 Phosphatidylcholine (PC) biosynthesis, PE => PC		**C,P**

M00092 Phosphatidylethanolamine (PE) biosynthesis, ethanolamine => PE **C,P**

M00093 Phosphatidylethanolamine (PE) biosynthesis, PA => PS => PE **C,P**

M00094 Ceramide biosynthesis **C**

M00066 Lactosylceramide biosynthesis **C**

M00099 Sphingosine biosynthesis **C**

M00100 Sphingosine degradation **C**

Nucleotide metabolism

Purine metabolism

M00048 Inosine monophosphate biosynthesis, PRPP + glutamine => IMP **C,P**

M00049 Adenine ribonucleotide biosynthesis, IMP => ADP,ATP **C,P**

M00050 Guanine ribonucleotide biosynthesis IMP => GDP,GTP **C,P**

M00546 Purine degradation, xanthine => urea **C,N**

Pyrimidine metabolism

M00051 Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP **C,P**

M00052 Pyrimidine ribonucleotide biosynthesis, UMP => UDP/UTP,CDP/CTP **C,P**

M00053 Pyrimidine deoxyribonuleotide biosynthesis, CDP/CTP => dCDP/dCTP,dTDP/dTTP **C,P**

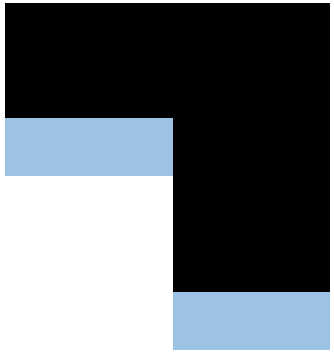
M00046 Pyrimidine degradation, uracil => beta-alanine, thymine => 3-aminoisobutanoate **C,P**

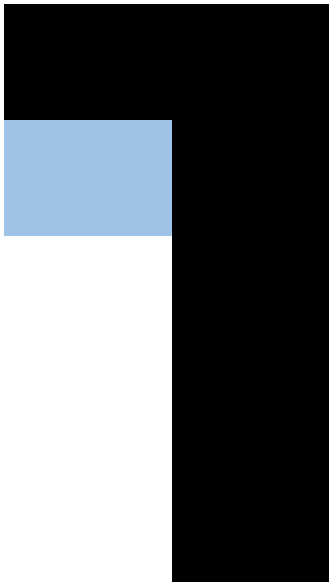
Amino acid metabolism

Serine and threonine metabolism

M00018 Threonine biosynthesis, aspartate => homoserine => threonine **C,N**

M00555 Betaine biosynthesis, choline => betaine **C,N**

M00020 Serine biosynthesis, glycerate-3P => serine		**C,N**
M00033 Ectoine biosynthesis, aspartate => ectoine		**C,N**
Cysteine and methionine metabolism		
M00021 Cysteine biosynthesis, serine => cysteine		**C,N**
M00338 Cysteine biosynthesis, homocysteine + serine => cysteine		**C,N**
M00034 Methionine salvage pathway		**C,N**
M00035 Methionine degradation		**C,N**
M00017 Methionine biosynthesis, apartate => homoserine => methionine		**C,N**
M00368 Ethylene biosynthesis, methionine => ethylene		**C,N**
Branched-chain amino acid metabolism		
M00019 Valine/isoleucine biosynthesis, pyruvate => valine / 2-oxobutanoate => isoleucine		**C,N**
M00535 Isoleucine biosynthesis, pyruvate => 2-oxobutanoate		**C,N**
M00570 Isoleucine biosynthesis, threonine => 2-oxobutanoate => isoleucine		**C,N**
M00432 Leucine biosynthesis, 2-oxoisovalerate => 2-oxoisocaproate		**C,N**
M00036 Leucine degradation, leucine => acetoacetate + acetyl-CoA		**C,N**
Lysine metabolism		
M00016 Lysine biosynthesis, succinyl-DAP pathway, aspartate => lysine		**C,N**
M00526 Lysine biosynthesis, DAP dehydrogenase pathway, aspartate => lysine		**C,N**
M00527 Lysine biosynthesis, DAP aminotransferase pathway, aspartate => lysine		**C,N**
M00433 Lysine biosynthesis, 2-oxoglutarate => 2-oxoadipate		**C,N**
Arginine and proline metabolism		

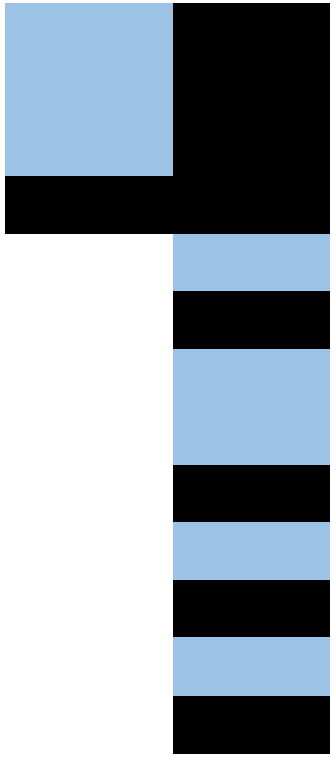
M00028 Ornithine biosynthesis, glutamate => ornithine		**C,N**
M00844 Arginine biosynthesis, ornithine => arginine		**C,N**
M00015 Proline biosynthesis, glutamate => proline		**C,N**
M00845 Arginine biosynthesis, glutamate => acetylcitrulline => arginine		**C,N**
M00047 Creatine pathway		**C,N**
Polyamine biosynthesis		
M00133 Polyamine biosynthesis, arginine => agmatine => putrescine => spermidine		**C,N**
M00134 Polyamine biosynthesis, arginine => ornithine => putrescine		**C,N**
M00136 GABA biosynthesis, prokaryotes, putrescine => GABA		**C,N**
Aromatic amino acid metabolism		
M00022 Shikimate pathway, phosphoenolpyruvate + erythrose-4P => chorismate		**C,N**
M00023 Tryptophan biosynthesis, chorismate => tryptophan		**C,N**
M00024 Phenylalanine biosynthesis, chorismate => phenylalanine		**C,N**
M00040 Tyrosine biosynthesis, prephanate => pretyrosine => tyrosine		**C,N**
M00025 Tyrosine biosynthesis, chorismate => tyrosine		**C,N**
M00042 Catecholamine biosynthesis, tyrosine => dopamine => noradrenaline => adrenaline		**C,N**
M00044 Tyrosine degradation, tyrosine => homogentisate		**C,N**
M00533 Homoprotocatechuate degradation, homoprotocatechuate => 2-oxohept-3-enedioate		**C,N**
M00545 Trans-cinnamate degradation, trans-cinnamate => acetyl-CoA		**C,N**
M00038 Tryptophan metabolism, tryptophan => kynurenine => 2-aminomuconate		**C,N**
Glycan metabolism		

Lipopolysaccharide metabolism

M00064 ADP-L-glycero-D-manno-heptose biosynthesis	**C**
M00072 N-glycosylation by oligosaccharyltransferase	**C**
M00073 N-glycan precursor trimming	**C**
M00056 O-glycan biosynthesis, mucin type core	**C**
M00070 Glycosphingolipid biosynthesis, lacto-series, LacCer => Lc4Cer	**C**
M00068 Glycosphingolipid biosynthesis, globo-series, LacCer => Gb4Cer	**C**
M00060 KDO2-lipid A biosynthesis, Raetz pathway, LpxL-LpxM type	**C**
M00866 KDO2-lipid A biosynthesis, Raetz pathway, non-LpxL-LpxM type	**C**
M00063 CMP-KDO biosynthesis	**C**
M00064 ADP-L-glycero-D-manno-heptose biosynthesis	**C**

Metabolism of cofactors and vitamins

Cofactor and vitamin metabolism	**C,P**
M00127 Thiamine biosynthesis, AIR => thiamine-P/thiamine-2P	**C,P**
M00125 Riboflavin biosynthesis, GTP => riboflavin/FMN/FAD	**C,P**
M00115 NAD biosynthesis, aspartate => NAD	**C,N**
M00119 Pantothenate biosynthesis, valine/L-aspartate => pantothenate	**C**
M00120 Coenzyme A biosynthesis, pantothenate => CoA	**C**
M00123 Biotin biosynthesis, pimeloyl-ACP/CoA => biotin	**C**
M00842 Tetrahydrobiopterin biosynthesis, GTP => BH4	**C**
M00843 L-threo-Tetrahydrobiopterin biosynthesis, GTP => L-threo-BH4	**C**



M00140 C1-unit interconversion, prokaryotes	**C**
M00121 Heme biosynthesis, plants and bacteria, glutamate => heme	**C, Fe**
M00846 Siroheme biosynthesis, glutamate => siroheme	**C,N,Fe**
M00112 Tocopherol/tocotorienol biosynthesis	**C**
M00124 Pyridoxal biosynthesis, erythrose-4P => pyridoxal-5P	**C,P**
M00572 Pimeloyl-ACP biosynthesis, BioC-BioH pathway, malonyl-ACP => pimeloyl-ACP	**C,P**
M00573 Biotin biosynthesis, BioI pathway, long-chain-acyl-ACP => pimeloyl-ACP => biotin	**C,P**
M00577 Biotin biosynthesis, BioW pathway, pimelate => pimeloyl-CoA => biotin	**C**
M00126 Tetrahydrofolate biosynthesis, GTP => THF	**C**
M00841 Tetrahydrofolate biosynthesis, mediated by PTPS, GTP => THF	**C**
M00122 Cobalamin biosynthesis, cobinamide => cobalamin	**C,Co**
M00117 Ubiquinone biosynthesis, prokaryotes, chorismate => ubiquinone	**C**
M00116 Menaquinone biosynthesis, chorismate => menaquinol	**C**
Biosynthesis of terpenoids and polyketides	
Terpenoid backbone biosynthesis	
M00096 C5 isoprenoid biosynthesis, non-mevalonate pathway	**C**
M00364 C10-C20 isoprenoid biosynthesis, bacteria	**C**
M00365 C10-C20 isoprenoid biosynthesis, archaea	**C**
M00095 C5 isoprenoid biosynthesis, mevalonate pathway	**C**
Xenobiotics biodegradation	
Aromatics degradation	

M00538 Toluene degradation, toluene => benzoate C

M00537 Xylene degradation, xylene => methylbenzoate C

M00551 Benzoate degradation, benzoate => catechol / methylbenzoate => methylcatechol lete) C

M00637 Anthranilate degradation, anthranilate => catechol C

M00568 Catechol ortho-cleavage, catechol => 3-oxoadipate C

M00569 Catechol meta-cleavage, catechol => acetyl-CoA / 4-methylcatechol => propanoyl-CoA C

M00540 Benzoate degradation, cyclohexanecarboxylic acid =>pimeloyl-CoA C

M00638 Salicylate degradation, salicylate => gentisate C

M00623 Phthalate degradation, phthalate => protocatechuate C

Drug resistance

M00627 beta-Lactam resistance, Bla system

M00745 Imipenem resistance, repression of porin OprD (13)

M00651 Vancomycin resistance, D-Ala-D-Lac type

M00726 Cationic antimicrobial peptide (CAMP) resistance, lysyl-phosphatidylglycerol (L-PG) synthase MprF

M00744 Cationic antimicrobial peptide (CAMP) resistance, protease PgtE

M00718 Multidrug resistance, efflux pump MexAB-OprM

M00642 Multidrug resistance, efflux pump MexJK-OprM

M00643 Multidrug resistance, efflux pump MexXY-OprM

M00769 Multidrug resistance, efflux pump MexPQ-OpmE

M00649 Multidrug resistance, efflux pump AdeABC

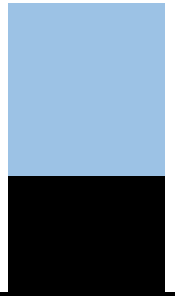
M00696 Multidrug resistance, efflux pump AcrEF-TolC

M00697 Multidrug resistance, efflux pump MdtEF-TolC

M00698 Multidrug resistance, efflux pump BpeEF-OprC

M00700 Multidrug resistance, efflux pump AbcA

M00714 Multidrug resistance, efflux pump QacA



13

14

15