

The global *Microcystis* interactome

Katherine V. Cook ^{1,2} Chuang Li ³ Haiyuan Cai ¹ Lee R. Krumholz,³ K. David Hambright ^{1,2*}
Hans W. Paerl ⁴ Morgan M. Steffen ⁵ Alan E. Wilson ⁶ Michele A. Burford ⁷ Hans-Peter Grossart ⁸
David P. Hamilton ^{7,9} Helong Jiang ¹⁰ Assaf Sukenik,¹¹ Delphine Latour,¹² Elisabeth I. Meyer ¹³
Judit Padisák ¹⁴ Boqiang Qin ¹⁰ Richard M. Zamor ^{15,a} Guangwei Zhu ¹⁰

¹Plankton Ecology and Limnology Laboratory, Department of Biology, The University of Oklahoma, Norman, Oklahoma

²Program in Ecology and Evolutionary Biology and the Geographical Ecology Group, Department of Biology, The University of Oklahoma, Norman, Oklahoma

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

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

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

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

⁷Australian Rivers Institute and School of Environment and Science, Griffith University, Nathan, Queensland, Australia

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

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

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

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

¹²Université Clermont Auvergne, CNRS, LMGE, Aubière Cedex, France

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

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

¹⁵Grand River Dam Authority, Vinita, Oklahoma

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 12 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 operational taxonomic units 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.

*Correspondence: dhambright@ou.edu

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional Supporting Information may be found in the online version of this article.

^aPresent address: Department of Natural Sciences, Northeastern State University, Tahlequah, Oklahoma.

Special Issue: Linking Metagenomics to Aquatic Microbial Ecology and Biogeochemical Cycles. Edited by: Hans-Peter Grossart, Ramon Massana, Katherine McMahon and David A. Walsh.

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) has resulted in the proliferation, intensification, and prolongation of cyanoHABs around the world (Carey et al. 2012; O’Neil 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, CO₂ 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 CO₂ and pH levels (Havens 2008), which are 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 toward 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

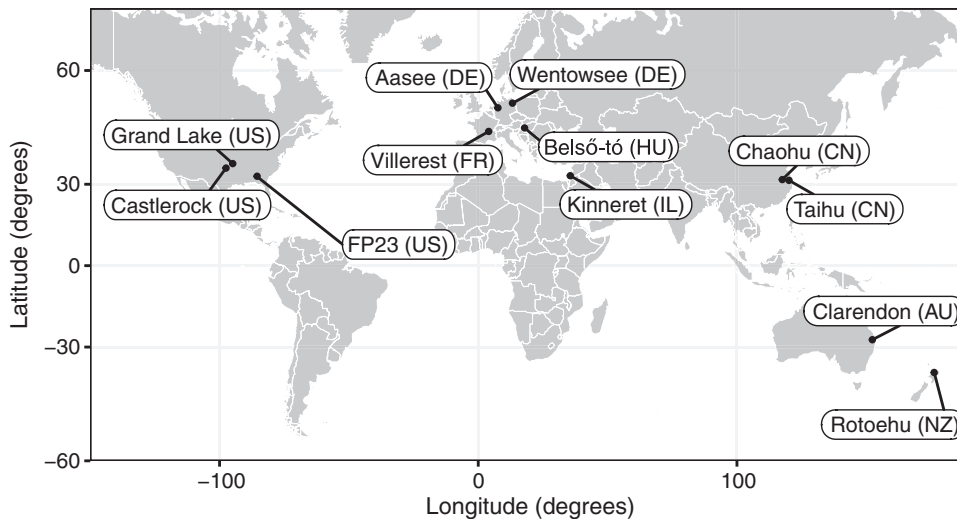


Fig. 1. Location of the 12 lakes across the globe. These samples represent a 280° longitudinal and 90° latitudinal gradient.

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). 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) 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) 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 μM , final concentration), and 25 μL PCR master mix containing DreamTaq polymerase (Thermo Fisher Scientific). Thermocycler conditions were as follows: 94°C for 3 min followed by 28 cycles of 94°C for 30 s, 55°C for 60 s, and 72°C for 75 s, 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) 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) 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) and requantified 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). OTUs 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 20 bp being discarded and removal of human genes (MetaWRAP; Uritskiy et al. 2018), the clean metagenomic data were 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. unpubl. preprint doi: arXiv:1312.4038v1 [q-bio.GN]), followed by integration using DAS Tool (using a threshold of $\geq 70\%$ 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 the Results section), 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 (10 *M. aeruginosa*, 1 *Microcystis viridis*, 1 *Microcystis panniformis*, 2 *Microcystis flos-aquae*, and 2 *Microcystis 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. 2016). 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.

Raw sequence data in this study have been deposited at the National Center for Biotechnology Information website under BioProject accession number PRJNA575023.

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). NTI 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 α NTI value less than -2 indicates taxa within the community are more distantly related than by chance (phylogenetically over-dispersed), a value greater than 2 indicates taxa are more closely related than expected by chance (phylogenetically clustered), and α NTI 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 α NTI values from across all communities was significantly different from that of the null distribution.

We used beta-mean-nearest-taxon-distance (β MNTD) and beta-nearest-taxon-index (β NTI) to quantify the phylogenetic distance among samples (see Swenson 2014 for R code). Similar to its α -diversity analog, β MNTD 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 β MNTD 999 times (Stegen et al. 2012). As before, β NTI is the number of standard deviations that the observed β MNTD is away from the mean of the null distribution. A value of β NTI less than -2 indicates more phylogenetic relatedness between samples than expected by chance, and a β NTI 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 β NTI 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) was 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.

Results

The sampled lakes spanned a 280° longitudinal and 90° latitudinal gradient ranging from 444 to 11,777 km apart and represented *M. aeruginosa* blooms from four continents (Fig. 1). After

removal of Eukaryota, singletons, chloroplasts, and mitochondrial reads, 3.25 million high-quality reads were represented by 454 unique (97% similarity) bacterial OTUs. *Microcystis* (927 total OTUs) was dominant in 10 of the 12 lakes and ranged from 65% to 84% of total sequence abundance. Both Taihu and Wentowsee communities were made up of less than 50% *Microcystis* (48% and 5%, respectively) sequences suggesting the lakes were not at peak bloom phase during sampling. As such, these two lakes were removed from subsequent analyses. Castlerock was also removed from subsequent analyses due to a loss of one of the triplicate samples. All *Microcystis* OTUs with greater than 1000 total abundance were reidentified to species using NCBI BLAST function (Altschul et al. 1990) and were all identified as *M. aeruginosa*, accounting for 79–85% of the total *Microcystis* across lakes and 53–68% of all OTU abundances across lakes. This confirms that, at least at the level of 16S rRNA sequences, *M. aeruginosa* is a cosmopolitan bloom-forming species.

The 454 non-cyanobacterial OTUs were associated with 35 bacterial classes, with most OTUs identified as Alphaproteobacteria, Bacteroidia, Gammaproteobacteria, Clostridia, Campylobacteria, Deltaproteobacteria, Negativicutes, Phycisphaerae, Gemmatimonadetes, Acidobacteriia, Ignavibacteria, SM1A07, Melainabacteria, Cytophagia, Parcubacteria, and Anaerolineae (Fig. 2b). The abundance and taxonomy of the associated bacteria differed among sites (Fig. 2b) with Alphaproteobacteria, Bacteroidia, and Gammaproteobacteria classes contributing most sequences across sites. Bray–Curtis dissimilarity values were relatively high 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,b; 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, $+\alpha$ NTI; $t = 10.13$, $p < 0.001$). Among lakes, the *Microcystis* microbiome communities also were significantly more phylogenetically related than expected by chance (Fig. 4, β NTI; $t = -12.65$, $p < 0.001$).

A total of 866 million metagenomic sequencing reads were generated from the nine lake *Microcystis*-microbiome communities. After trimming and filtering, 612 million (52.4–137.6 million per lake) clean reads were generated (Supplementary Table S1).

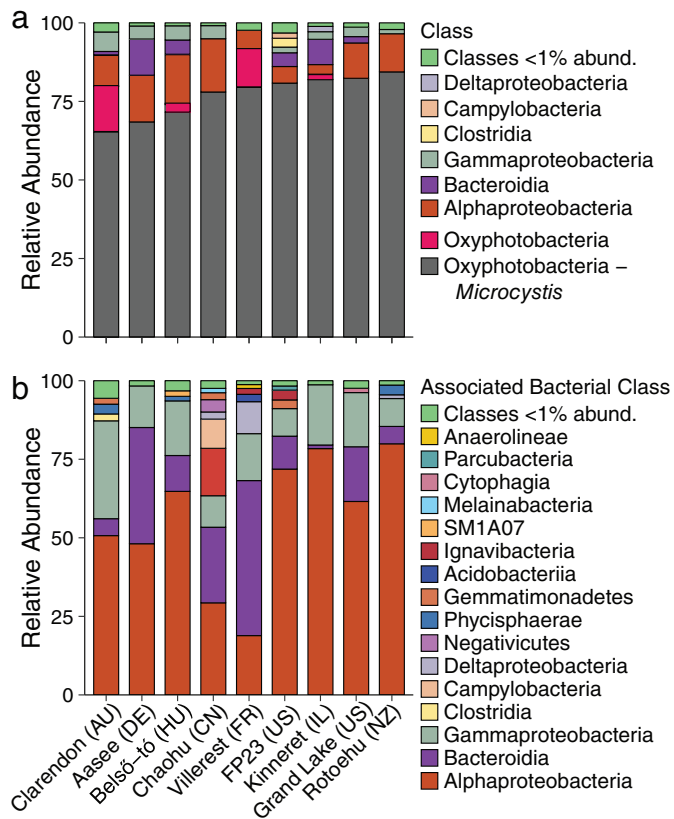


Fig. 2. (a) Relative abundance of Bacteria classes or subclasses in the nine lakes. The lakes are arranged in order from left to right of increasing percent *Microcystis* in the community. Classes less than 1% of the total relative abundance were grouped together as a single group denoted “<1% abund.” 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.

Most of the clean metagenomic reads (55.2–76.6% 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 KEGG orthology. By contrast, after removing duplicate and rRNA genes, analysis of the total metagenomic data produced 407,658 and 54,312 protein-coding genes for the microbiome and *Microcystis*, respectively, of which 139,384 (more than twice the number compared with MAGs) and 17,817 (3629 more), respectively, were successfully mapped to the KEGG orthology. Microbiome gene diversity was similar across lakes (Fig. 5; low Bray–Curtis values) and did not differ with geographic distance between the lakes (Fig. 5; BC GLM DE = 1.97%, $p = 0.41$).

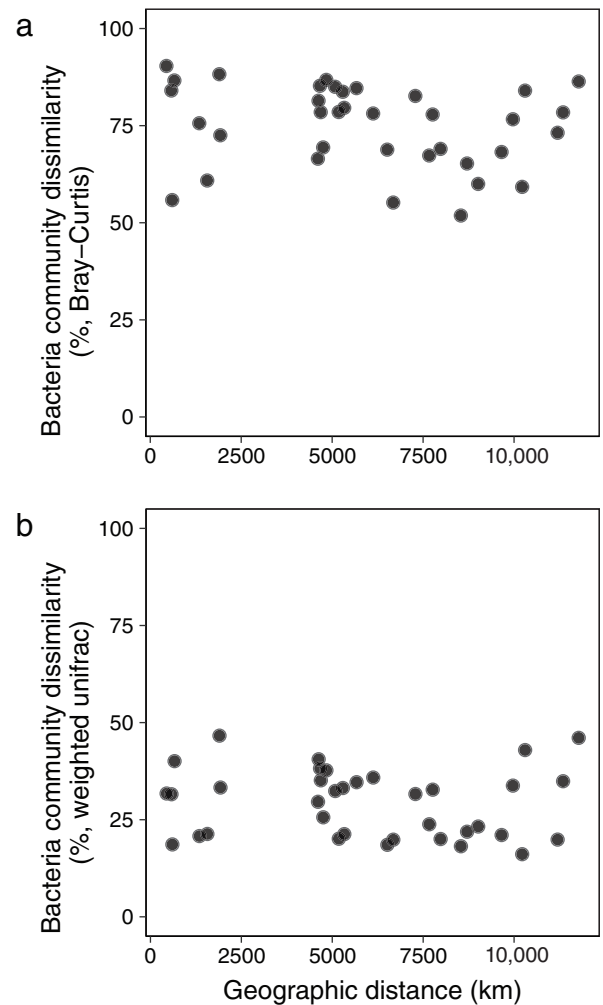


Fig. 3. Scatter plots of community dissimilarity in the microbiome as related to geographic distance. (a) The nonsignificant (GLM deviance explained = 3.7%, $p = 0.2$) relationship between taxonomic Bray–Curtis dissimilarity and geographic distance where the higher Bray–Curtis values indicate fewer species 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.

Shared pathways were mostly assigned to carbohydrate metabolism; amino acid, nucleotide, and fatty acid biosynthesis; and cofactor and vitamin biosynthesis (Fig. 6). The microbiome bacteria contained unique pathways for organic carbon transportation and degradation and vitamin B₁₂ synthesis not found in *Microcystis*. Numerous pathways for organic carbon degradation (D-galacturonate, D-glucuronate, galactose, glycogen, fatty acid, purine, pyrimidine) and transportation (e.g., proline, maltose, galactose, maltose, mannose, D-xylose, fructose, rhamnose, glycerol), as well as complete pathways for degradation of aromatics (mostly anthropogenic pollutants, such as toluene, xylene, benzene, phthalate) were also identified in the microbiome. Numerous important anaerobic bacterial pathways, including nitrogen fixation, denitrification, and dissimilatory

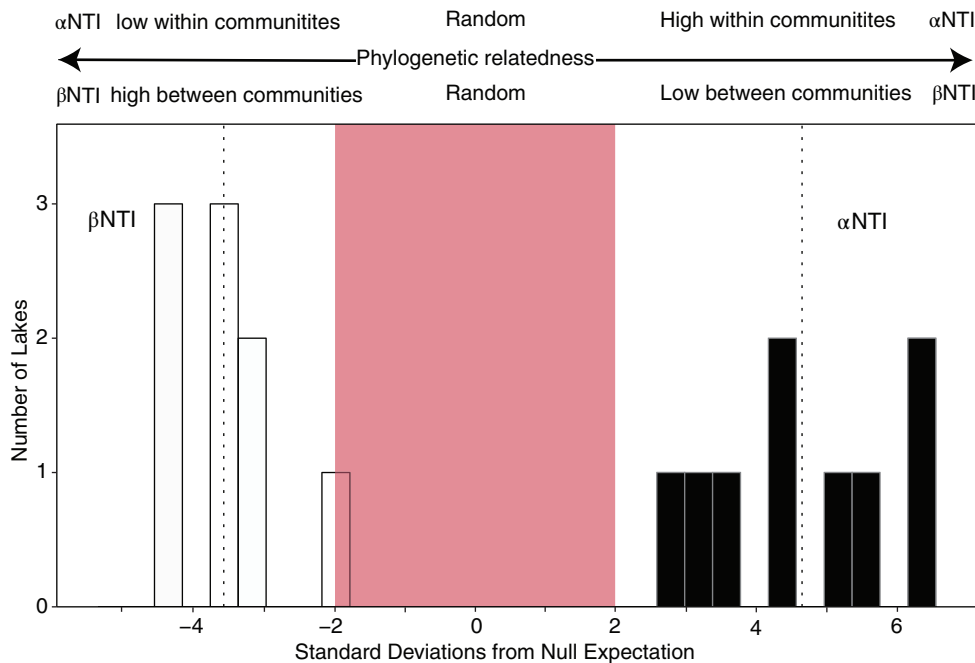


Fig. 4. Distributions of *within* community phylogenetic relatedness (α NTI, nearest taxon index), and the phylogenetic relatedness *between* two communities (β NTI) of the nine sampled lakes. Values below -2 or above $+2$ SD from the null (indicated by the red rectangle) are statistically significantly different from random. Black dashed lines indicate the mean of the observed distributions. The mean of the α NTI distribution is 4.64 and the mean of the β NTI distribution is -3.58 . α NTI 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 α NTI was significantly different from the null ($t = 10.13$, $p < 0.001$). β NTI 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 β NTI is significantly different from random ($t = -12.65$, $p < 0.001$).

sulfate reduction were also detected in the *Microcystis* microbiome, suggesting the anaerobic bacteria play an important role in nutrient cycling during *Microcystis* blooms in these

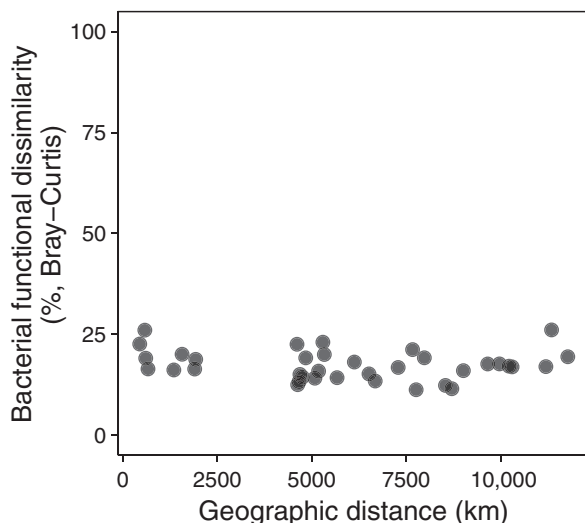


Fig. 5. The dissimilarity between the *Microcystis* microbiome community's metagenomic function was not significantly correlated with geographic distance (GLM DE = 1.97% , $p = 0.41$) and was overall low (low values of Bray-Curtis dissimilarity).

lakes. With the exception of some genes related to carbon cycling (purine, pyrimidine, salicylate, and catechol degradation), as well as bacterial pathways involved in methane metabolism (methane oxidation and formaldehyde assimilation), the majority of potential biochemical function appeared to be associated with the 43 microbiome MAGs (Fig. 6 and Table S2).

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 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

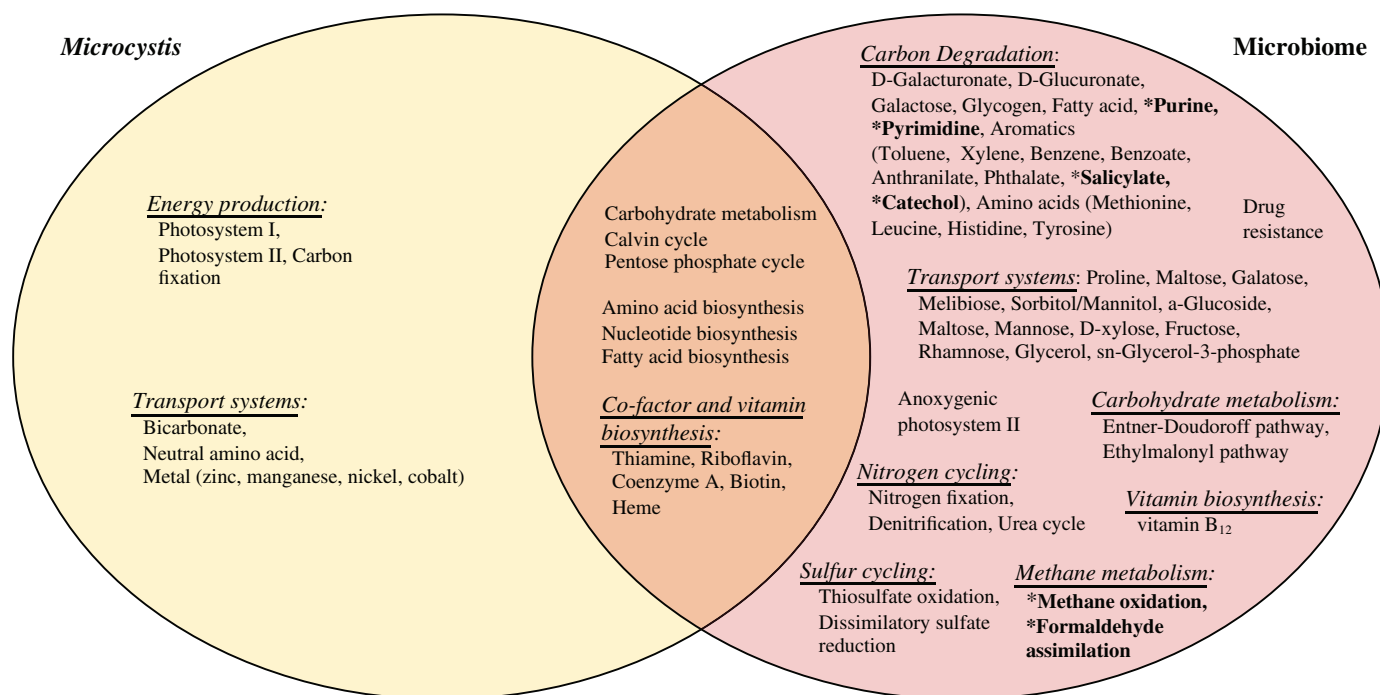


Fig. 6. Venn diagram showing the distribution of complete or nearly complete (no more than one gene missing) KEGG modules in *Microcystis* and the microbiome bacteria. See Supplementary Table S2 for details and indication for involvement in major elemental cycling. KEGG modules in bold print with asterisks were detected in the full metagenome data but not in the *Microcystis* or the microbiome bacterial MAGs.

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.

M. 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, b), indicating no distance-decay relationship as would be expected for dispersal-limited species (Nekola and White 1999; Green and Bohannan 2006; Nemergut et al. 2013). We also conclude that the functional potential of microbial communities is more highly conserved than their taxonomic composition (Figs. 3a, 5). Similarly, Steffen et al. (2012) found that the cyanobacterial bloom-associated bacterial communities across three lakes (Erie, Taihu, St. Marys) differed taxonomically, while being functionally similar. Thus, while OTU identity was variable across *Microcystis* microbiome bacteria, functional potential appears to have been quite similar providing evidence that under bloom

conditions, *M. aeruginosa* is accompanied by a common suite of bacterial functionality, potentially forming an interactome.

We found that *Microcystis* microbiome bacteria have functional potential not found in *Microcystis*, and functional similarity was preserved globally. Pathways involving photosynthesis (excluding the anoxygenic photosystem II pathway) and carbon fixation were only detected in *Microcystis* (Fig. 6). The microbiome bacteria could potentially be contributing to carbon recycling within the aggregates as many of the pathways found only in the bacteria are related to carbohydrate breakdown (e.g., d-galacturonate, galactose, and glycogen degradation) and transport (e.g., numerous transport systems). The microbiome bacteria are likely tightly associated with the carbon sources found in the *Microcystis* aggregates. Bacteria associated with phytoplankton blooms have been found to utilize the carbon source glycolate from phytoplankton (Lau et al. 2007; Paver and Kent 2010) and contain the glycolate oxidation pathway. Unfortunately, the KEGG database does not contain this gene or related pathways thus we were unable to detect it in our samples. Every metagenomic function database has its shortcomings, and for future studies we would recommend using multiple databases to cover as 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 microcystin 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. unpubl preprint, <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 B₁₂ 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 B₁₂ and other essential vitamins. Indeed, previous studies have also suggested that vitamin B₁₂ was provided to *Microcystis* by the associated bacteria and, moreover, that this relationship was mutually beneficial for both groups (Xie et al. 2016; Li et al. 2018).

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

Loss of necessary metabolic functions is more common in bacterial communities than previously thought (Hottes et al. 2013). Morris et al. (2012) found that the marine cyanobacterium, *Prochlorococcus*, has lost many oxidative-stress genes and instead relies on other microbes for removal of hydrogen

peroxide from the microenvironment. As such, *Prochlorococcus* may benefit from hydrogen peroxide removal via the actions of “helper” microbes in the form of a “leaky” public good (sensu Morris et al. 2012), and the smaller genome of *Prochlorococcus* affords it a selective advantage. A recent genomic study has shown that *M. aeruginosa* also has a reduced genome with remarkable redundancy consisting of a set of ~ 2400 core genes and a large, variable pangenome, an additional set of genes unique to different *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 and Harold 1952; Bell and Mitchell 1972; Paerl and Gallucci 1985). We saw through β NTI analysis that *Microcystis* microbiome communities were significantly phylogenetically related (Fig. 4), with the negative value of β NTI indicating a common environmental filter across the lakes—*Microcystis* (sensu Dini-Andreote et al. 2015). Similarly, multiple studies (Yang et al. 2017; Shi et al. 2018; Xu 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; Shi et al. 2018; Xu 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 $\geq 99\%$, 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 have 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. aeruginosa* and its microbiome. We conclude that the coevolution within an interactome, that is, between *M. aeruginosa* and its microbiome, could help explain the global distribution of, and dominance by, *M. aeruginosa* in diverse freshwater ecosystems.

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. doi:10.1089/cmb.2012.0021
- Batista, A., J. Woodhouse, H.-P. Grossart, and A. Giani. 2018. Methanogenic archaea associated to *Microcystis* sp. in field samples and in culture. *Hydrobiologia* **831**: 1–10. doi:10.1007/s10750-018-3655-3
- 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
- Bourne, D. G., P. G. Dennis, S. Uthicke, R. M. Soo, G. W. Tyson, and N. Webster. 2013. Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. *ISME J.* **7**: 1452–1458. doi:10.1038/ismej.2012.172
- Burke, C., P. Steinberg, D. Rusch, S. Kjelleberg, and T. Thomas. 2011. Bacterial community assembly based on functional genes rather than species. *PNAS* **108**: 14288–14293. doi:10.1073/pnas.1101591108
- Cai, H., H. Jiang, L. R. Krumholz, and Z. Yang. 2014. Bacterial community composition of size-fractionated aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater lake. *PLoS One* **9**: e102879. doi:10.1371/journal.pone.0102879
- Caporaso, J. G., K. Bittinger, F. D. Bushman, T. Z. DeSantis, G. L. Andersen, and R. Knight. 2009. PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–267. doi:10.1093/bioinformatics/btp636
- Caporaso, J. G., and others. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**: 335–336. doi:10.1038/nmeth.f.303
- Carey, C. C., B. W. Ibelings, E. P. Hoffmann, D. P. Hamilton, and J. D. Brookes. 2012. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. *Water Res.* **46**: 1394–1407. doi:10.1016/j.watres.2011.12.016
- Carmichael, W. W. 2001. Health effects of toxin-producing cyanobacteria: “The CyanoHABs”. *Hum. Ecol. Risk Assess.* **7**: 1393–1407. doi:10.1080/20018091095087
- Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot, and G. E. Likens. 2009. Controlling eutrophication: Nitrogen and phosphorus. *Science* **323**: 1014–1015. doi:10.1126/science.1167755

- Croft, M. T., A. D. Lawrence, E. Raux-Deery, M. J. Warren, and A. G. Smith. 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**: 90–93. doi:[10.1038/nature04056](https://doi.org/10.1038/nature04056)
- Croft, M. T., M. J. Warren, and A. G. Smith. 2006. Algae need their vitamins. *Eukaryot. Cell* **5**: 1175–1183. doi:[10.1128/EC.00097-06](https://doi.org/10.1128/EC.00097-06)
- Dini-Andreote, F., J. C. Stegen, J. D. van Elsas, and J. F. Salles. 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *PNAS* **112**: E1326–E1332. doi:[10.1073/pnas.1414261112](https://doi.org/10.1073/pnas.1414261112)
- Dziallas, C., and H. Grossart. 2011. Temperature and biotic factors influence bacterial communities associated with the cyanobacterium *Microcystis* sp. *Environ. Microbiol.* **13**: 1632–1641. doi:[10.1111/j.1462-2920.2011.02479.x](https://doi.org/10.1111/j.1462-2920.2011.02479.x)
- Dziallas, C., and H.-P. Grossart. 2012. Microbial interactions with the cyanobacterium *Microcystis aeruginosa* and their dependence on temperature. *Mar. Biol.* **159**: 2389–2398. doi:[10.1007/s00227-012-1927-4](https://doi.org/10.1007/s00227-012-1927-4)
- Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461. doi:[10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461)
- Edgar, R. C. 2013. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**: 996–998. doi:[10.1038/nmeth.2604](https://doi.org/10.1038/nmeth.2604)
- Eiler, A., and S. Bertilsson. 2004. Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environ. Microbiol.* **6**: 1228–1243. doi:[10.1111/j.1462-2920.2004.00657.x](https://doi.org/10.1111/j.1462-2920.2004.00657.x)
- Erb, T. J., I. A. Berg, V. Brecht, M. Muller, G. Fuchs, and B. E. Alber. 2007. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway. *PNAS* **104**: 10631–10636. doi:[10.1073/pnas.0702791104](https://doi.org/10.1073/pnas.0702791104)
- Fierer, N. 2017. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**: 579–590. doi:[10.1038/nrmicro.2017.87](https://doi.org/10.1038/nrmicro.2017.87)
- Fogg, G. E., and P. W. Harold. 1952. The production of extracellular nitrogenous substances by a blue-green alga. *Proc. R. Soc. B* **139**: 372–397. doi:[10.1098/rspb.1952.0019](https://doi.org/10.1098/rspb.1952.0019)
- Garcia, S. L., M. Buck, K. D. McMahon, H.-P. Grossart, A. Eiler, and F. Warnecke. 2015. Auxotrophy and intrapopulation complementarity in the “interactome” of a cultivated freshwater model community. *Mol. Ecol.* **24**: 4449–4459. doi:[10.1111/mec.13319](https://doi.org/10.1111/mec.13319)
- Giovannoni, S. J., J. Cameron Thrash, and B. Temperton. 2014. Implications of streamlining theory for microbial ecology. *ISME J.* **8**: 1553–1565. doi:[10.1038/ismej.2014.60](https://doi.org/10.1038/ismej.2014.60)
- Green, J., and B. J. M. Bohannan. 2006. Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* **21**: 501–507. doi:[10.1016/j.tree.2006.06.012](https://doi.org/10.1016/j.tree.2006.06.012)
- Gregory, T. R., and others. 2007. Eukaryotic genome size databases. *Nucleic Acids Res.* **35**: D332–D338. doi:[10.1093/nar/gkl828](https://doi.org/10.1093/nar/gkl828)
- Häder, D.-P., and K. Gao. 2015. Interactions of anthropogenic stress factors on marine phytoplankton. *Front. Environ. Sci.* **3**:14. doi:[10.3389/fenvs.2015.00014](https://doi.org/10.3389/fenvs.2015.00014)
- Hahn, M. W., J. Jezberová, U. Koll, T. Saueressig-Beck, and J. Schmidt. 2016. Complete ecological isolation and cryptic diversity in *Polynucleobacter* bacteria not resolved by 16S rRNA gene sequences. *ISME J.* **10**: 1642–1655. doi:[10.1038/ismej.2015.237](https://doi.org/10.1038/ismej.2015.237)
- Harke, M. J., M. M. Steffen, C. J. Gobler, T. G. Otten, S. W. Wilhelm, S. A. Wood, and H. W. 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](https://doi.org/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*. New York: Springer. doi:[10.1007/978-0-387-75865-7_33](https://doi.org/10.1007/978-0-387-75865-7_33)
- 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](https://doi.org/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](https://doi.org/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**: 471–483. doi:[10.1038/s41579-018-0040-1](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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. doi:[10.7717/peerj.1165](https://doi.org/10.7717/peerj.1165)
- 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. doi:[10.1093/bioinformatics/btq166](https://doi.org/10.1093/bioinformatics/btq166)
- 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](https://doi.org/10.1093/nar/gks808)

- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359. 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. doi:10.1128/AEM.01965-06
- 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. *ISME J.* **11**: 1813–1824. doi:10.1038/ismej.2017.49
- 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**: 746. doi:10.3389/fmicb.2018.00746
- MacArthur, R. H. 1984. Geographical ecology: Patterns in the distribution of species. Princeton Univ. 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–e00012. doi:10.1128/mBio.00036-12
- 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 N₂ 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. Microbiol.* **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
- Paerl, H. W., and others. 2016. It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environ. Sci. Technol.* **50**: 10805–10813. doi:10.1021/acs.est.6b02575
- Paerl, H. W., H. Xu, M. J. McCarthy, G. Zhu, B. Qin, Y. Li, and W. S. Gardner. 2011. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Res.* **45**: 1973–1983. doi:10.1016/j.watres.2010.09.018
- Pande, S., and C. Kost. 2017. Bacterial unculturability and the formation of intercellular metabolic networks. *Trends Microbiol.* **25**: 349–361. doi:10.1016/j.tim.2017.02.015
- Parveen, B., V. Ravet, C. Djediat, I. Mary, C. Quiblier, D. Debroas, and J.-F. Humbert. 2013. Bacterial communities associated with *Microcystis* colonies differ from free-living communities living in the same ecosystem: Bacterial diversity inside *Microcystis* colonies. *Environ. Microbiol. Rep.* **5**: 716–724. doi:10.1111/1758-2229.12071
- Paver, S. F., and A. D. Kent. 2010. Temporal patterns in glycolate-utilizing bacterial community composition correlate with phytoplankton population dynamics in humic lakes. *Microb. Ecol.* **60**: 406–418. doi:10.1007/s00248-010-9722-6

- Pérez, A. A., D. A. Rodionov, and D. A. Bryant. 2016. Identification and regulation of genes for cobalamin transport in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* **198**: 2753–2761. doi:10.1128/JB.00476-16
- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490. doi:10.1371/journal.pone.0009490
- 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, D1
- R Development Core Team. 2018. 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**: 836–843. 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 two continents. *PLoS One* **7**: e44002. doi:10.1371/journal.pone.0044002
- Stegen, J. C., X. Lin, A. E. Konopka, and J. K. Fredrickson. 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6**: 1653–1664. doi:10.1038/ismej.2012.22
- Steiner, K., S. A. Wood, J. Puddick, I. Hawes, D. R. Dietrich, and D. P. Hamilton. 2017. A comparison of bacterial community structure, activity and microcystins associated with formation and breakdown of a cyanobacterial scum. *Aquat. Microb. Ecol.* **80**: 243–256. doi:10.3354/ame01852
- Swan, B. K., and others. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *PNAS* **110**: 11463–11468. doi:10.1073/pnas.1304246110
- Swenson, N. G. 2014. Functional and phylogenetic ecology in R. Springer. doi:10.1007/978-1-4614-9542-0
- Uritskiy, G. V., J. DiRuggiero, and J. Taylor. 2018. MetaWRAP—A flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**: 158. doi:10.1186/s40168-018-0541-1
- Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**: 5261–5267. doi:10.1128/AEM.00062-07
- Wawrik, B., and others. 2012. Field and laboratory studies on the bioconversion of coal to methane in the San Juan Basin. *FEMS Microbiol. Ecol.* **81**: 26–42. doi:10.1111/j.1574-6941.2011.01272.x
- Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: An example for rain Forest trees. *Am. Nat.* **156**: 145–155. doi:10.1086/303378
- Wienhausen, G., B. E. Noriega-Ortega, J. Niggemann, T. Dittmar, and M. Simon. 2017. The exometabolome of two model strains of the *Roseobacter* group: A marketplace of microbial metabolites. *Front. Microbiol.* **8**: 1985. doi:10.3389/fmicb.2017.01985
- Wu, Y.-W., B. A. Simmons, and S. W. Singer. 2015. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607. doi:10.1093/bioinformatics/btv638
- Xie, M., M. Ren, C. Yang, H. Yi, Z. Li, T. Li, and J. Zhao. 2016. Metagenomic analysis reveals symbiotic relationship among bacteria in *Microcystis*-dominated community. *Front. Microbiol.* **7**: 56. doi:10.3389/fmicb.2016.00056
- Xu, H., and others. 2018. Contrasting network features between free-living and particle-attached bacterial communities in Taihu Lake. *Microb. Ecol.* **76**: 303–313. doi:10.1007/s00248-017-1131-7
- Yang, C., and others. 2017. Distinct network interactions in particle-associated and free-living bacterial communities during a *Microcystis aeruginosa* bloom in a Plateau Lake. *Front. Microbiol.* **8**: 1202. doi:10.3389/fmicb.2017.01202

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 Fig. 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-1P01ES028939-01 to H.W.P., NSF DEB 1831106 to M.M.S., and through an NSF Graduate Research Fellowship to K.V.C. This

Cook et al.

Microcystis interactome

work will be submitted by K.V.C. as partial fulfillment of the requirements for a Ph.D. degree from the University of Oklahoma.

Submitted 21 February 2019

Revised 22 July 2019

Accepted 24 September 2019

Conflict of Interest

None declared.

Associate editor: Katherine McMahon