Lake Erie Field Trials to Advance Autonomous Monitoring of Cyanobacterial Harmful Algal Blooms

Paul A. Den Uyl, Luke R. Thompson, Reagan M. Errera, James M. Birch,Christina M. Preston, William Ussler III, Colleen E. Yancey, Subba Rao Chaganti, Steven A. Ruberg, Gregory J. Doucette, Gregory J. Dick, Christopher A. Scholin, and Kelly D. Goodwin

Supplementary Material

## Sampling details

In 2018, water samples were collected by the LRAUV- 3G ESP from a total of 18 waypoints across western Lake Erie (WLE) starting initially with a series of seven locations, in order of collection: 1) Stony Point, 2) WE04, 3) between WE02 and WE04, 4) WE12, 5) WE02, 6) between WE02 and WE08, and 7) WE08 (**Figure 1**, **Table S1** and **Table S2** for details). Cartridges were, in general, used in a triplicate configuration by sequentially engaging archive, lysis, and archive cartridges at each station. After instrument retrieval for battery recharge, collections using the triplicate sampling scheme continued at five additional waypoints in order of: 8) WE04, 9) between WE02 and WE04, 10) WE02, 11) WE08, and 12) West-northwest of WE04. Following these waypoints, only archive cartridges remained, which were subsequently used at six waypoints in order of: 13) WE04, 14) between WE04 and WE08, 15) between WE04 and WE13, 16) WE13, 17) east of and in between WE13 and WE16, and 18) East-northeast of WE16 to exhaust the sampling capacity of the 3G ESP. In 2019, water was collected by the LRAUV-3G ESP at 20 waypoints in an area generally outlined by stations WE02, 2G ESP, WE12, WE13, WE04 and WE08 (**Figure 1**, see **Table S1** and **Table S2** for details).

Most of the matched samples were collected around the 2G ESP (16 samples total, 8 matched sample pairs, **Table S1**). The 2G ESP, a predecessor to the 3G ESP, is substantially larger than the 3G ESP and is deployed to autonomously monitor the cyanotoxin microcystin in WLE from a stationary, moored location. Matched samples were also collected between stations WE04 and WE08 (4 samples, 2 matched sample pairs, **Table S1**).

Chlorophyll *a* measurements collected from monitoring stations accessible to the LRAUV-3G ESP were analyzed during three dates in 2018 (8/20, 8/27, and 9/4) and two dates in 2019 (8/19, and 8/28) marking times either the day before (8/20/2018), the day after (9/4/2018, 8/28/2019), or during (8/27/18, 8/19/19) autonomous deployments. Each date includes sample measurements from six of eight routine monitoring stations (WE02, 04, 08, 12, 13, and 16). WE06 and WE09 (**Figure 1**) were not included, because they are generally too shallow to be accessed by the LRAUV-3G ESP.

## Positive and negative LRAUV-3G ESP controls

An archive cartridge positive control was run in 2018 consisting of a lake water sample from the prior weekly monitoring cruise known to contain high cyanobacterial biomass. The source of control water was selected by determining the station sample with the highest visible biomass. In total, 23 mL of sample water was processed through the archive cartridge before the LRAUV was sealed (“canned”) for deployment (**Table S1**). Negative controls for the archive cartridges consisted of PCR grade water (Invitrogen, Carlsbad, CA, USA). In 2018, negative controls were processed through an archive cartridge before deployment (n=1), in between Leg I and II (n=2), and after deployment (n=1). In 2019, negative controls were processed prior to (n=1) and after (n=1) the deployment.

## Quantitative real-time PCR (qPCR) assays

Total Cyanobacteria and Toxin Gene qPCR assays (Phytoxigene, Inc.) were run according to the manufacturer’s instructions. These kits are self-contained in that they include internal amplification controls (IAC) in the Total Cyanobacteria assay and have accompanying standard curves from the manufacture (Phytoxigene, Inc.). IAC are used to assess for sample inhibition. Inhibition is identified when an IAC of an individual sample is more than 1.5 cycles above the IAC of the no-template control (NTC). An NTC is a sample included in the same run with template DNA replaced with molecular-grade water. All reactions utilized 5 µL of template or molecular-grade water (NTC) for negative controls. To maintain values within the range of the standard curves, DNA extracted from the environmental samples was first diluted 1:5, with the exception of the 2018 DNA samples analyzed by the Toxin Gene assay which were processed undiluted. Detailed qPCR results are provided in Table S7.

All 2018 runs were performed on an Applied Biosystems 7500 Fast Real-Time PCR System, and all 2019 runs were performed in an Applied Biosystems QuantStudio 6 Flex Real-Time PCR System. The same Phytoxigene standards were run on both instruments and instrument bias was not observed. Consistent readings also were demonstrated by testing an environmental sample on both instruments (undiluted and diluted 1:5, each in duplicate) for both qPCR assays (data not shown).

## Amplicon sequencing

Unlike the Earth Microbiome Project protocol, which currently contains barcodes on the forward primer 515F to enable the usage of various reverse primer constructs to obtain longer amplicons, the custom NEXTFlex kit’s barcodes are located on the reverse primer (Quince et al. 2011, Parada et al. 2015) (**Table S3**). Barcoded 16S and 18S-ITS amplicons were pooled together and sequenced using the Illumina MiSeq v3 600 cycle (2 x 300 bp) kit on the same sequencing run with 20% phiX spike-in, as recommended by the NEXTFlex kit manufacturer (BiooScientific). For both 2018 and 2019, DNA and standards were processed following the same protocol with one exception. The exception was that pooled 18S ITS amplicons were amplified with 18S ITS Amplicon-Seq Kit (NOVA-4210-05) containing unique single-indexes in 2018, but in 2019 they contained unique dual-indexed barcodes. This study reports amplicon sequencing results from the 16S rRNA gene amplicon sequence (16S amplicon) target only.

## DNA Extraction Efficiency

DNA extraction efficiency was characterized by laboratory recoveries of the microcystin producing culture *Microcystis aeruginosa* LE3 (Brittain et al. 2000) and non-microcystin-producing *M. aeruginosa* 2386 obtained from UTEX The Culture Collection of Algae (University of Texas at Austin, Austin, TX). Each culture was extracted in duplicate and processed using the Total Cyanobacteria assay to determine sample concentration (gc/μL) using the methods detailed in the main text. Operating under the assumption that each strain of *Microcystis* possessed two copies of the 16S rRNA gene in its genome (Smith et al. 2021, Yancey et al. 2022), the laboratory recoveries for *M. aeruginosa* LE3 (microcystin-producing) ranged from 159% to 255% and *M. aeruginosa* 2386 (non-microcystin producing) strains ranged from 94% to 166%.

## Negative qPCR controls

The limit of detection for the Total Cyanobacteria and Toxin Gene assays is 40 copies per reaction (cpr), and the limit of quantification for each assay is 100 cpr according to the assay manufacturer. qPCR results for NTC were below the limit of detection for both Total Cyanobacteria and Toxin Gene qPCR assays. Results for extraction blanks (see main text) and archive negative controls processed on the LRAUV-3G ESP (see above) were also below the limit of detection for Toxin Gene qPCR assays. Total Cyanobacteria qPCR assay tests on extraction blanks and archive negative controls were often above the limit of quantification; however, they had a substantially lower final concentration (max. 1,226 copies per mL) than the lowest environmental sample processed by the LRAUV-3G ESP (21,071 copies per mL).

## Positive sequencing controls

Two types of mock community positive controls were included as separate samples during amplicon sequencing on the MiSeq platform (**Table S8**). Results from 16S amplicon sequencing of Community Standard D6306 (Zymo Research, Irvine, CA, USA) were consistent with the expected composition for all three dilutions tested in 2019. In 2018, results were as expected for a single sample of standard D6306, although a low level of contamination with an AlphaproteobacteriaASV was observed (0.01% abundance). Results from Community Standard D6311 (Zymo Research, Irvine, CA, USA), which contained staggered concentrations of DNA, were consistent with the expected composition for those taxa present at 0.012% or greater in the mock community. *Enterococcus faecalis* and *Staphylococcus aureus*, with respective theoretical abundances of 0.00067% and 0.0001%, were not detected in either the 2018 or 2019 sequencing runs, suggesting an overall detection limit near 0.01%. Community Standard D6311 showed *Microcystis aeruginosa* contamination (0.02% abundance) in 2018, no contamination above the impurity level (<0.01%) was observed in 2019.

## Size fraction comparisons

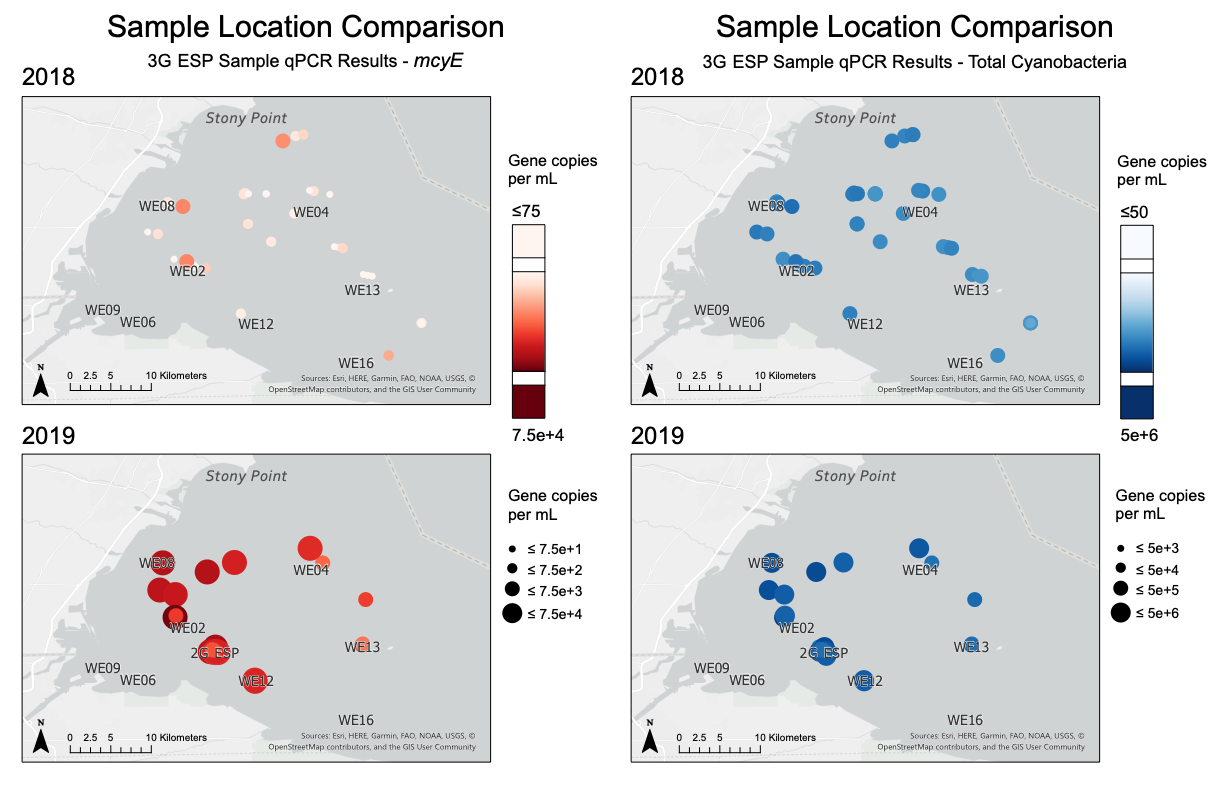
Principal component analysis (PCA) was used to compare the microbial communities obtained from the two stacked filters in the archive cartridges for both 16S amplicon and metagenomic sequencing. As expected, the 5-μm size fraction clustered separately from the 0.22-μm fraction in the PCA plots for both the amplicon and metagenomic sequencing data (**Figure S8**). PERMANOVA analysis also showed that beta diversity differed significantly between the two size fractions for both sequencing methods (α= 0.05) (**Table S7**). By offering the 0.22-μm size fraction data, this instrument provided microbiome characterization that was not available from routine monitoring.

During the 2018 deployment, some archive cartridges failed to fully evacuate reagent during the 2018 deployment, allowing liquid from the 5-μm size fraction to potentially contact the 0.2 μm filter during removal of filters from sample pucks. Subsequent software engineering modifications resolved the issue, and reagents were adequately evacuated from all samples during the second year’s field mission. This experience illustrates the process of field demonstrations informing iterative engineering design during technology development.

# Supplementary Figures

# Diagram Description automatically generated

**Figure S1**. Study location with outline of sample collection and processing steps.

****

**Figure S2**. qPCR results (gc/mL) for the Toxin Gene (left) and Total Cyanobacteria (gc/mL) assays mapped by location for 5-μm samples collected in 2018 (top) and 2019 (bottom). Samples shown here were collected using LRAUV-3G ESP archive cartridge filters only, with results coded by symbol color and size.



**Figure S3**. Relationship between *mcyE* and total cyanobacteria samples collected by the LRAUV-3G ESP in a) 2018 and b) 2019.



**Figure S4**. Correlation between Chlorophyll Index (CI, mean value within a 5-km radius of each collection point) and the Total Cyanobacteria qPCR assay result (log gc/ml).

**Map

Description automatically generated**

**Figure S5**. (LEFT) Cyanobacterial Index (CI) from nearest assigned value of processed satellite imagery on exampled dates, gray represents cloud cover; (RIGHT) 3G ESP sample locations and total cyanobacteria (gc/mL) results from collections on corresponding date.

****

**Figure S6**. Box plot of Chlorophyll *a* concentrations (µg/L) obtained from routine WLE monitoring vessel operations during the approximate timeframe of LRAUV deployments.

**Calendar

Description automatically generated with medium confidence**

**Figure S7**. Microcystis ASV sequence alignment. Bold and underlined nucleotides indicate the bases used for ASV naming scheme. Red (highlighted) nucleotides show the position where ASV Microcystis CT(A)G differs.

****

**Figure S8**. PCA plots for 16S amplicon (ASV counts) and metagenomic (species count profiles) sample mates, with size fractions analyzed together. Community composition was significantly different between size fractions for both sequencing methods.

References

Brittain, S. M., J. Wang, L. Babcock-Jackson, W. W. Carmichael, K. L. Rinehart and D. A. Culver (2000). "Isolation and Characterization of Microcystins, Cyclic Heptapeptide Hepatotoxins from a Lake Erie Strain of Microcystis aeruginosa." Journal of Great Lakes Research **26**(3): 241-249. doi:<https://doi.org/10.1016/S0380-1330(00)70690-3>

Parada, A. E., D. M. Needham and J. A. Fuhrman (2015). "Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples." Environmental Microbiology **18**(5): 1403-1414. doi:10.1111/1462-2920.13023

Quince, C., A. Lanzen, R. J. Davenport and P. J. Turnbaugh (2011). "Removing noise from pyrosequenced amplicons." BMC bioinformatics **12**: 38-38. doi:10.1186/1471-2105-12-38

Smith, D. J., J. Y. Tan, M. A. Powers, X. N. Lin, T. W. Davis and G. J. Dick (2021). "Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time." Environmental Microbiology. doi:10.1111/1462-2920.15514

Yancey, C. E., D. J. Smith, P. A. D. Uyl, O. G. Mohamed, F. Yu, S. A. Ruberg, et al. (2022). "Metagenomic and Metatranscriptomic Insights into Population Diversity of <i>Microcystis</i> Blooms: Spatial and Temporal Dynamics of <i>mcy</i> Genotypes, Including a Partial Operon That Can Be Abundant and Expressed." Applied and Environmental Microbiology **0**(0): e02464-02421. doi:doi:10.1128/aem.02464-21